河北省软科学计划项目（10457204D-13）
河北省社会科学基金项目（HB2011QR44） 资助



基于数据挖掘技术的新农村建设调查分析

李春林 屈驳韵 申 博
武依明 梁 艳 刘金凤 著

清华大学出版社

基于数据挖掘技术的新农村建设调查分析

李春林　屈驳韵　申　博
武依明　梁　艳　刘金凤　著



清华大学出版社
北　京

内 容 简 介

本书作者在河北省软科学研究计划项目《河北省科技进步、农村能源开发与新农村建设关系研究》(项目编号：10457204D-13)、河北省社会科学基金项目《基于第六次全国人口普查数据的统计分析和数据挖掘》(项目编号：HB2011QR44)的支持下，通过设计新农村建设调查问卷和农村养老满意度调查问卷，深入农村开展实地调查，了解农民对新农村建设及农业生产、农民生活的满意度，基于决策树、关联规则等数据挖掘模型和面板数据模型，深入探索农村现代化、新农村建设、能源使用和农民满意度的内在联系，全面系统地分析了科技进步、新农村建设和能源开发之间的关系，探讨了农村养老满意度的影响因素和养老模式的改革，提出了具有针对性的、可操作的对策建议。

图书在版编目 CIP 数据

基于数据挖掘技术的新农村建设调查分析/李春林等著. --北京：清华大学出版社，2014
ISBN 978-7-302-38590-5

Ⅰ. ①基…　Ⅱ. ①李…　Ⅲ. ①数据采集—计算机应用—农村—社会主义建设—调查研究—河北省　Ⅳ. ①F327.22-39

中国版本图书馆 CIP 数据核字(2014)第 273651 号

责任编辑：陈　明　赵从棉
封面设计：
责任校对：赵丽敏
责任印制：

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮编**：100084
社 总 机：010-62770175　　**邮购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 刷 者：
装 订 者：
经　　销：全国新华书店
开　　本：148mm×210mm　**印张**：6.125　**插页**：3　**字数**：180 千字
版　　次：2014 年 12 月第 1 版　　**印次**：2014 年 12 月第 1 次印刷
印　　数：1～　000
定　　价：　.00 元

产品编号：054184-01

前　言

根据我国现代化建设三步走战略目标，到 2050 年前后，赶上世界中等发达国家水平，基本实现现代化，已经成为全国人民的热切期盼。要实现中国现代化，就必须先实现农村工业化、城镇化，进一步实现农村的现代化，农村现代化是整个中国现代化的重要组成部分。近年来，河北省农村经济快速发展，人民生活水平不断提高，农村改革深入推进，农村经济体制、发展方式和社会结构发生了深刻变革。但是，随着农村工业化和城镇化的加速推进，农民的精神和物质需求会发生巨大的变化，农村对能源的需求将持续大幅度增长，能源与农村资源、环境和社会发展的矛盾也会日益突出。加上人口老龄化的来临，使得农业生产、农村建设和农民生活之间的矛盾进一步加剧。为此，我们在利用河北省第二次农业普查数据对农民的能源使用状况进行分析的基础上，在河北省软科学研究计划项目《河北省科技进步、农村能源开发与新农村建设关系研究》（项目编号：10457204D-13）、河北省社会科学基金项目《基于第六次全国人口普查数据的统计分析和数据挖掘》（项目编号：HB2011QR44）的支持下，通过设计新农村建设调查问卷和农村养老满意度调查问卷，深入农村开展实地调查，了解农民对新农村建设及农业生产、农民生活的满意度，探索农村现代化、新农村建设、能源使用和农民满意度的内在联系，全面系统地分析了科技进步、新农村建设和能源开发之间的关系，探讨了农村养老满意度的影响因素和养老模式的改革，提出了具有针对性的、可操作的对策建议。

研究结果显示：第一，耐用消费品的使用是农民生活能源消费增加的主要原因，机械化生产逐渐成为农业生产的主要方式，生产用电能、水能需求日益增加，资金缺口加大。第二，通过对新农村建设试点村和普通村的对比调查及对新农村建设调查数据的决策树分析和关联

规则分析发现，新农村建设作为改善农民生产、生活条件的一个惠民举措，受到了农民的拥护和肯定。但随着农民文化程度和素质的提高，农民的期望值在增加，其满意度在下降。影响农民满意度的主要因素依次为收入、社会风气、居住环境和领导干部问题，农民最关注的是对收入起决定作用的农村经济发展问题，特别是农村经济发展的核心问题——农村的科技进步和人才培养问题。第三，利用决策树的C5.0、CART算法和面板模型分析了农民生活满意度的影响因素及生产现代化、生活现代化和农民素质现代化对农民生活满意度的影响机理。①生产方面，畜牧业等对第一产业的补贴政策以及种植业等基本情况对农民生活满意度影响最大，其次农业机械化水平对农民生活满意度起正面作用，二、三产业收入占农民收入比重提高，进而影响农民生活满意度。②生活方面，农村基础设施建设以及城乡商品市场情况对农民收入满意度影响最强；其次为社会保障制度，含最低生活保障、医疗保障和养老保险；第三位的为农民收入来源对农民生活满意度的影响。③农民素质方面，村民对教育环境和教育条件的满意度、提高农民素质组织的培训班的评价影响生活满意度，此外农民反应希望能够获得文化知识和技术培训。第四，经济来源是否独立、居住状况、低保制度的实施情况的好坏和是否参加农村合作医疗是影响老年人养老满意度的关键因素。

在上述分析结论的基础上，我们提出了相应的"提升农民生活水平，实现'三农'协调发展"的对策建议如下：

第一，在农村能源利用和发展方面，从农村实际情况出发，着力提高农村煤炭的利用效率；加快全省农村的电源、电网建设；大力发展农村沼气产业，建设节约型新农村；全面发展农村可再生新能源，实现农村多能互补的能源模式；依靠政府指导和政策调节，逐步改善农村能源的使用状况。

第二，在新农村建设方面，加大力度进行新农村建设和新民居建设，特别是通过新型城镇化建设，盘活农村宅基地资源，充分发挥城镇化的聚集效应，促使农村产业发展，增加农民收入；加大科技投入，促进农民素质提高；在新农村建设和新民居建设中做到统筹兼顾，因地制

宜，政府主导，返利于民，提高农民的政策理解度。

第三，提高农村现代化水平，提升农民生活满意度。①推进农业生产现代化，促进农村繁荣，为新农村建设提供强有力的产业支撑，推进农村二、三产业结构调整和发展方式转变；加快变革传统农业的生产方式，使之产业化和现代化，才能逐村改善经济状况，提高农民生活水平，改善贫困状况；改变传统农业以粮食生产为主的单一结构，实现农、林、牧、渔等方面多层次发展。②推进农村生活现代化，加大对农村基础设施建设的投入；在初次分配注重效率，通过政策干预加强政府对收入再分配的调节职能，建立健全社会保障体系，为经济发展和社会稳定提供良好的环境；大力发展农村卫生事业；加强农村环境综合整治。③推进农民素质现代化，要加强农民思想宣传教育，要继续普及九年制义务教育和农村职业教育，大力实施科教兴农战略，实施农村基层公共文化服务体系建设工程，繁荣农村文化，形成农村倡导健康文明的新风尚。

第四，通过促进农村经济发展，建设农村老年宜居社区；加强老年家庭建设，支持居家养老；促进老龄产业发展，建设多元化养老机构；完善农村社会保障和医疗保障制度，促进政策实施等措施，积极调动个人、家庭、社区和国家的力量，建立一个以居家为基础、社区为依托、机构为支撑，农村社会养老保障制度和医疗保障制度完善的农村养老模式，为我国广大农村地区老年人营造一个适合的养老环境，实现“老有所养、老有所医、老有所教、老有所学、老有所为、老有所乐”的目标。

目　　录

第 1 篇　河北省农村能源使用状况分析

第 2 篇　加快新农村建设　提高农民满意度

第 3 篇 实现农村现代化 提升农民生活质量

第 4 篇 适应人口老龄化需求 加快农村养老模式改革

第 1 篇

河北省农村能源使用状况分析

第 1 章　农村能源的概念和范畴

1.1　能源的概念和范畴

能源是指人类取得能量的来源，包括已开采出来可供使用的自然资源与经过加工转换的能量的来源。对能源可以进行如下分类，见表 1-1。

表 1-1　能源分类表

<table>
<tr><th colspan="3">类　别</th><th>可再生能源</th><th>不可再生能源</th></tr>
<tr><td rowspan="3">一次能源</td><td rowspan="2">常规能源</td><td>商品能源</td><td>水力(大型)
核能(增殖堆)
地热</td><td>化石燃料(煤、油、天然气)
核能</td></tr>
<tr><td>传统能源(非商品能源)</td><td>生物质能(薪柴、秸秆、人畜粪便)
太阳能(自然干燥等)
水力(水车等)
风力(风车、风帆等)
畜力</td><td></td></tr>
<tr><td>非常规能源</td><td>新能源</td><td>生物质能(燃料作物所制沼气、酒精等)
太阳能(收集电、发光池)
水力(小水电)
风力(风力机等)
海洋能
地热</td><td></td></tr>
<tr><td>二次能源</td><td colspan="4">电力、煤炭、沼气、汽油、煤油、重油等油制品，蒸汽、热水、压缩空气、氢能等</td></tr>
</table>

1. 可再生能源与不可再生能源

在自然界中可以不断再生并有规律地得到补充的能源,称为可再生能源。经过亿万年形成的、短期内无法恢复的能源,称为不可再生能源,它随着大规模的开采使用,储量越来越少,总有枯竭之日。

2. 常规能源与非常规能源(新能源)

在一定历史时期和科学技术水平下,已经被人们广泛应用的能源,称为常规能源。许多古老但由于诸多限制未得到广泛利用的传统能源,如太阳能、风能、生物质能等,以及用新发展的先进技术利用的能源,如氢能等,称为新能源。

3. 商品能源与非商品能源(传统能源)

以经济流通领域中的地位加以区分,商品能源是指进入市场进行买卖的能源,如煤炭、石油及其制品、焦炭、电力等。非商品能源是指那些一般不通过市场的能源,如某些传统能源,秸秆、薪柴、人畜粪便等。虽然它们有时在当地市场上也有买卖,但规模很小,也未将其列入正式商品,称为非商品能源。

4. 一次能源与二次能源

自然界现成存在、可直接取得而又不改变其基本形态的能源称为一次能源,或称初级能源。由一次能源经过加工转换成另一种形态的能源产品称为二次能源,也称次级能源。

1.2 农村能源的范畴

以上能源的范畴是从世界范围来说的,而农村能源,这个专用名词多见于城乡差别明显的发展中国家,在发达国家由于农村和城市已无较大差别,因此只有农业能源的概念。在我国,农村能源包括两方面的含义,一方面是指能在农村就地开发利用的能源资料,包括生物质能(作物秸秆、人畜粪便、薪柴以及沼气等)、小水电、太阳能、风能和地热能等;另一方面是指农村的生产和生活用能。从目前农村用能构成来看,60%仍是农村就地开发利用的农村能源。商品常规能源,如煤、

油、电、气等，约占 40%，主要用在生产上。能源是现代社会发展的动力，我国农村正处在一个伟大的历史转变时期，由于农业生产、乡镇企业、交通运输、物资供给、生活水平等的提高都需要能源，因此农村能源的合理建设和科学发展已成为农村经济发展的当务之急。尽管水资源不属于能源范畴，但是水资源与能源一样是农民日常生活、生产不可缺少的关键要素。同时，因为河北省是水资源严重短缺的省份，多年平均年降水量 540mm，居全国第 23 位；多年平均地表水资源 152 亿 m^3，居全国第 25 位；人均、亩均水量不足全国平均数的 1/8 和 1/10；地表水连年减少、地下水严重超采，导致地面沉降、环境恶化、地方病流行(尤其以氟中毒最为严重)，水已经成为河北省社会、经济发展的主要制约因素。再生水资源也是水电站发电的直接动力。因此此次研究大胆拓宽能源的范畴，将农民对水资源的使用状况与能源使用状况一起进行分析。

第2章　河北省农民能源使用状况总体特征研究

农村能源消费状况的规模和结构、数量与质量、静态与动态等多方面的综合分析、评价和判断，是我们认识和研究河北省农民能源消费总体特征的基本方法。从河北省第二次农业普查结果可以看出全省农民能源使用主要是在以下几个方面通过用能的家庭户数得到直接或间接的反映：首先是生活用能，如主要炊事用能、辅助炊事用能、取暖用能；其次是生产用能，如灌溉用水、大牲畜数量；再次是基于深层次的间接用能，如各种电器、生产机械、交通工具等用能。研究结果表明，河北省农民能源使用的总体特征概括如下。

2.1　生活用能状况及特征分析

1. 柴草和煤是全省农村的主导炊事能源

普查数据显示（见表2-1），河北省农村居民炊事使用的各种能源中，柴草和煤是全省农民使用的主导炊事能源。全省1375万户农村常住户中，从主要炊事用能源的使用户数看，有669.9万户，占48.72%的农户主要使用柴草；有611.75万户，占44.48%的农户主要使用煤；有82.98万户，占6.03%的农户主要使用煤气或天然气；有6.01万户，占0.44%的农户主要使用沼气；有4.16万户，占0.3%的农户主要使用电；有0.0224万户，仅占0.002%的农户使用太阳能；使用其他能源的有0.37万户，仅占0.025%。

表 2-1　农村炊事能源使用常住户数比例表　　%

指　　标	主要炊事用能源农户比重	辅助炊事用能源农户比重
柴草	48.72	18.51
煤	44.48	24.87
煤气、天然气	6.03	38.01
沼气	0.44	1.75
电	0.30	15.68
太阳能	0.002	0.05
其他	0.025	1.14

2. 火炕和暖气是农民取暖的主要方式

农民生活用能的另一重要用途是取暖。取暖的方式主要有火炕、暖气、空调等，火炕主要以燃烧柴草为主，暖气主要以燃烧煤为主，空调则以耗电为主。因此，从取暖方式的户数可以反映出农民能源的使用状况。普查数据显示（见图 2-1），全省 1375 万户农村常住户中，有 606.92 万户，占 44.13%的农户使用火炕取暖；有 485.08 万户，占 35.27%的农户使用暖气取暖；有 14.33 万户，占 1.04%的农户用空调取暖；还有 31.47 万户，占 2.29%的农户没有取暖设施。这说明河北省农民主要以使用柴草和煤等能源来取暖。

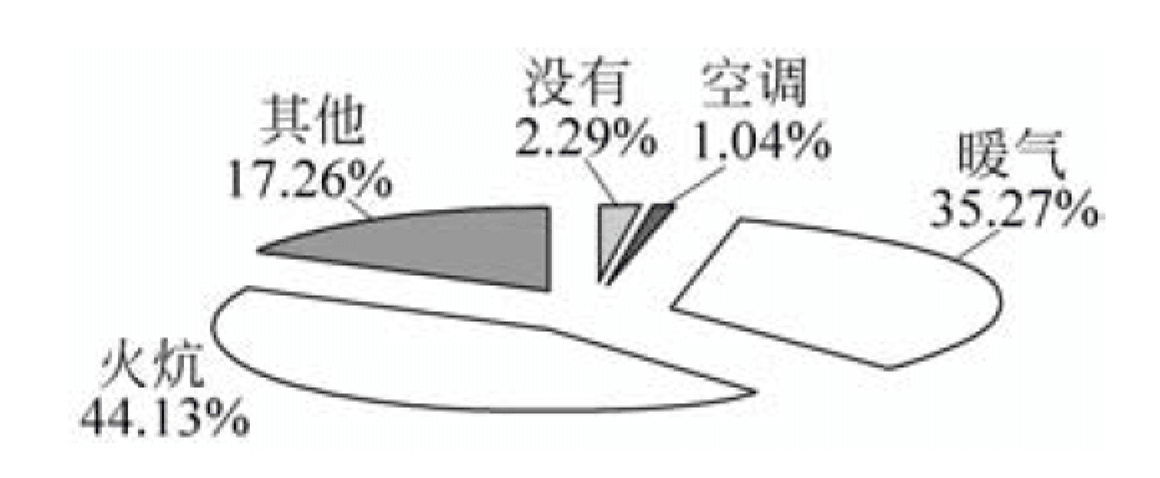


图 2-1　全省农民取暖方式及其比重图

3. 深井水是农民最主要的饮用水源

普查结果显示（见图 2-2），河北省农民饮用水资源还是相当健康的，全省 1375 万户农村常住户中，有 940.13 万户，占 68.4%的农户有条件使用入户管道水。从饮用水资源的情况看，有占 84.89%的农户饮用井水，其中有 1019.86 万户，占 74.16%的农户饮用深井水；有

147.62 万户，占 10.73％的农户饮用浅井水。另有 187.5 万户，占 13.63 的农户饮用净化处理的水。其余以江河湖水、池塘水、雨水等为饮用水的农户仅占 1.09 个百分点。但同时全省仍有 64.50 万户，占 4.7％的农户存在获取饮用水困难的情况。

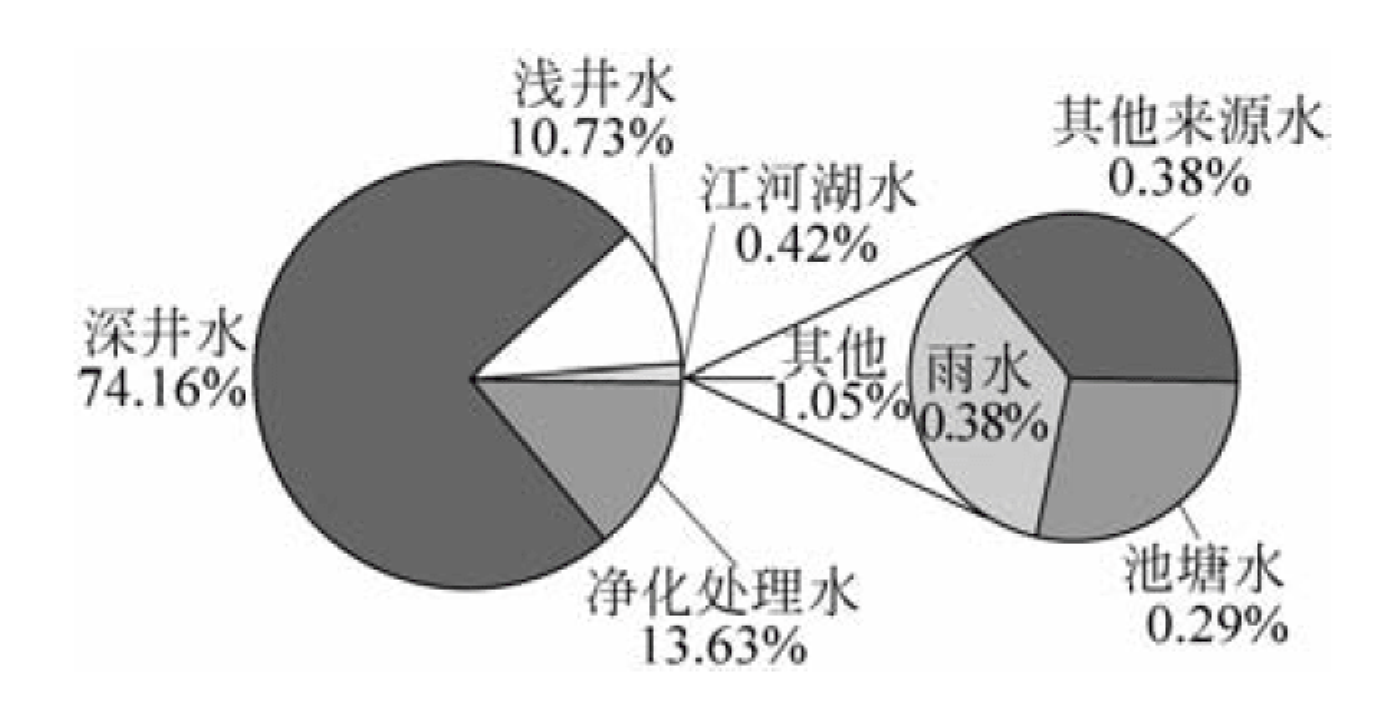


图 2-2　全省农户饮用水资源及其比重图

4. 耐用消费品使用将成为农民生活能源消费的重要原因之一

统计数据显示，随着新农村建设的不断深入，河北省农民的收入也在不断增加，2000 年全省农民人均纯收入为 2478.86 元，2006 年为 3801.82 元，2012 年已达到 8081.40 元，按可比价格计算，2006 年比 2000 年增加了 60.77 个百分点，而 2012 年比 2006 年增加了 112.57％。随着收入的不断增加，农民对高耗能的耐用消费品需求也越来越大(见图 2-3)，其中电冰箱 2006 年的百户拥有量为 32.62 台，2012 年达到 82.93 台；空调器 2006 年的百户拥有量为 5.45 台，2012 年已达到 41.62 台；家用计算机 2006 年的百户拥有量为 2.02 台，2012 年已达到 30.40 台。同时，移动电话机、彩色电视机、洗衣机等耐用品的拥有量也非常大。因此消耗在耐用消费品上的能源也越来越多，耐用消费品将日益成为农民能源消费的重要原因之一。

综上可知，河北省目前能保障绝大部分农民的基本生活用能。但全省农民生活能源消费结构单一，柴草和煤是保障河北省农民基本生活正常进行的主导能源。而尤其像太阳能、沼气等新能源在农民生活使用能源中所占比例很小。由参考文献可知秸秆、薪柴的热转换效率为 19％，煤炭为 25％，这说明河北省目前农民生活能源使用的特点是

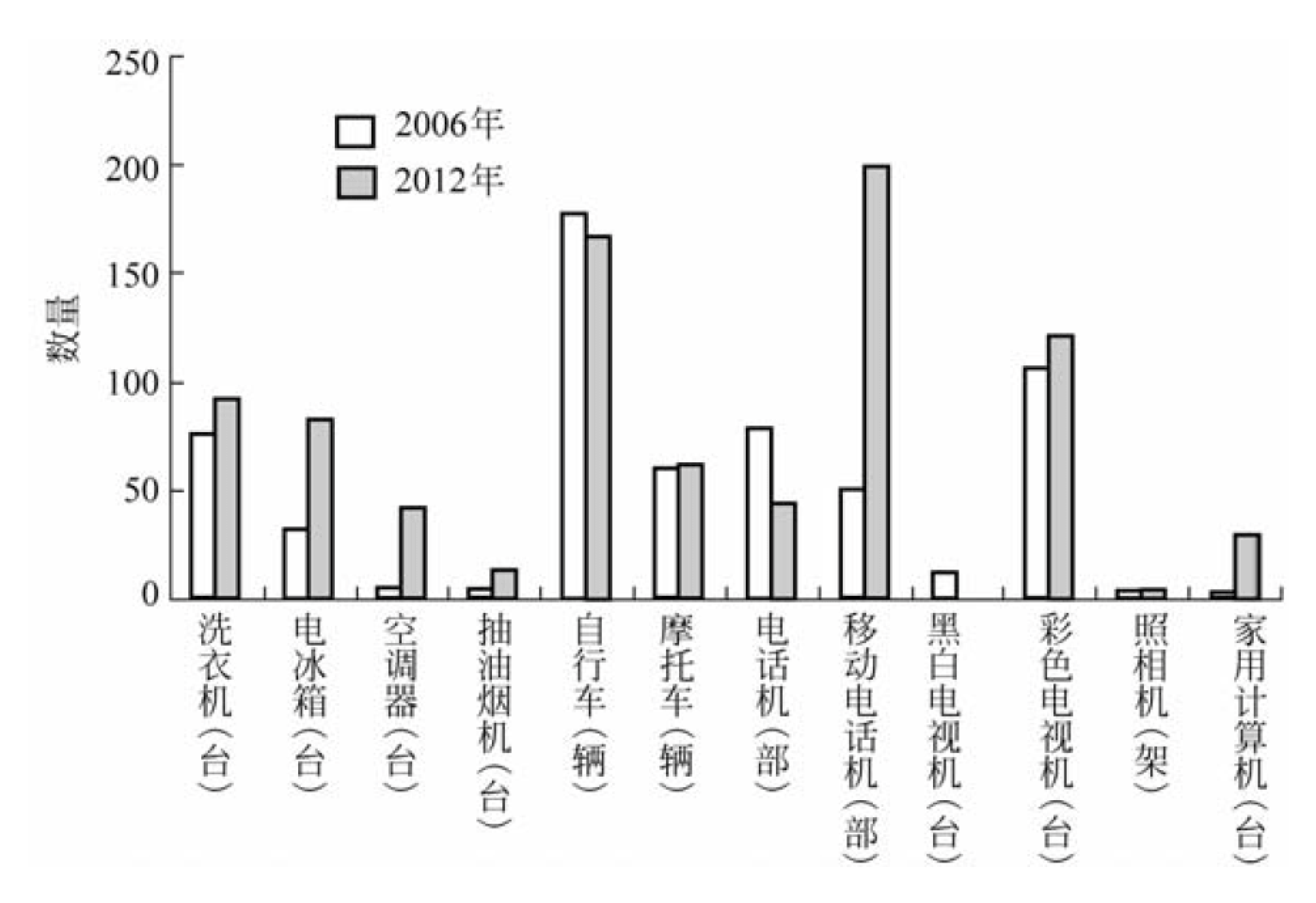


图 2-3 每百户农民平均拥有耐用消费品数量图

高能耗,低效率,高污染。《河北省十一五能源发展规划》指出:“河北省是环境污染比较严重的地区,且目前全省煤炭、天然气等供应缺口依然很大。”这在一定程度上加重了河北省资源利用的负担,影响了河北省能源安全的健康发展。

2.2 生产用能状况及特征分析

研究农村能源基础设施建设状况,可以从另一个方面反映出农民能源使用的规模和结构。

1. 农村生产用电量逐年增加

由图 2-4 可以看出,2004 年以后河北省农村用电量直线攀升,2012 年农村用电总量已达到 593.94 亿 kW·h,和 2003 年的 216.77 亿 kW·h 相比,用电量净增长 174.00%,2003—2012 年年均递增增长率为 11.85%。从全省电力设施建设来看,全省乡、村及村以下办水电站数和装机容量虽有波动,但基本呈现逐年增加的趋势,但增加缓慢,无法适

应用电量持续增加的需要。到了2011年和2012年,全省乡、村及村以下办水电站数和装机容量才有了一个飞速的发展,2011年全省乡、村及村以下办水电站数和装机容量239个和37.94亿kW,与2010年135个和6.32亿kW相比,分别增加了77.04个百分点和500.32个百分点。这说明无论是河北省乡、村及村以下办水电站数量还是装机容量,在2010年以前电力设施建设欠账较多,但最近几年在还清欠账的基础上,逐渐建立了一个适合农村经济发展的电能基础。

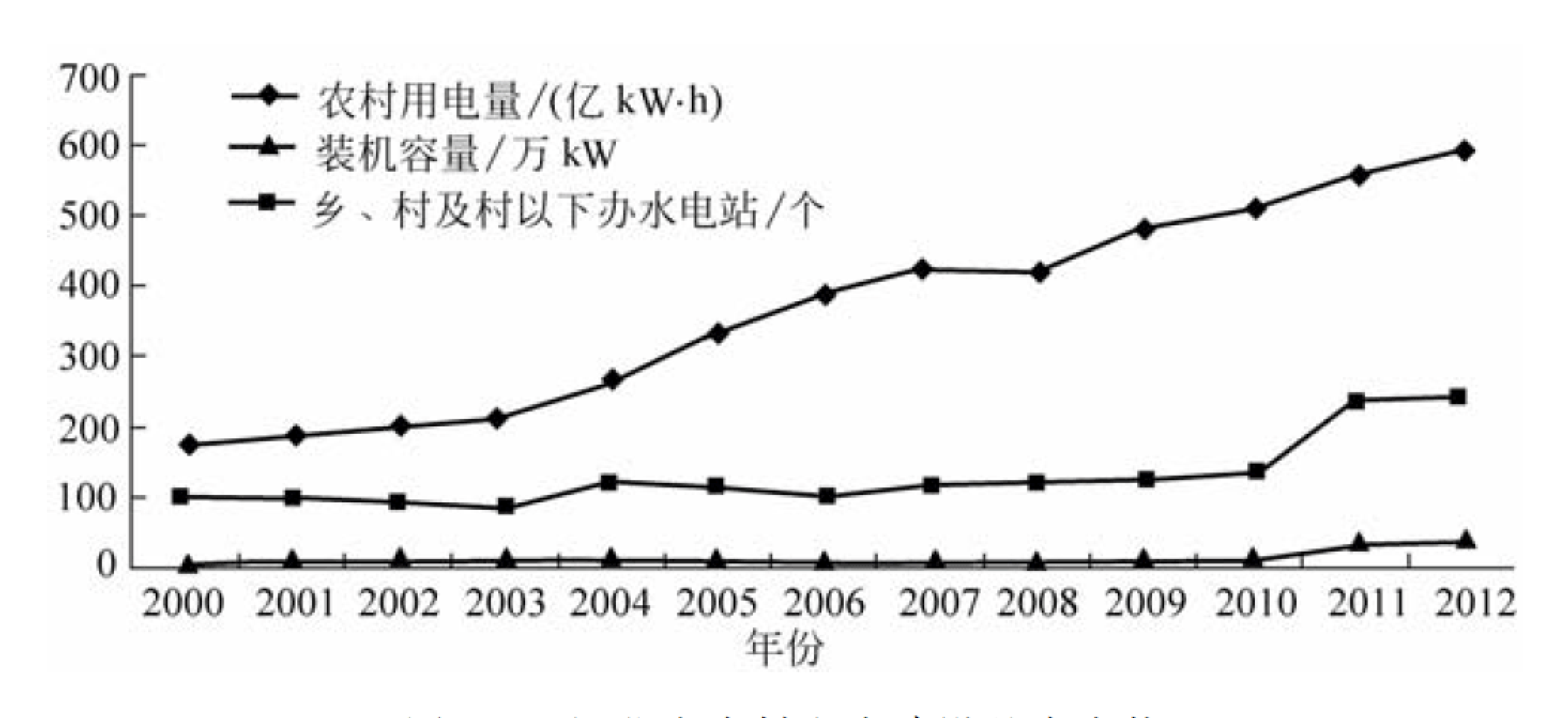


图 2-4　河北省农村电力建设基本走势

2. 农村水能建设需求大,资金少

河北省目前农作物总播种面积为8 777 370公顷,其中农田有效灌溉面积为4 569 775公顷,这说明加强对农村水利设施建设将有利于加快全省新农村建设的步伐。普查数据显示(见表2-2),河北省49 248个村中有42 942个,占87.2%的村在正常年景下用水有保障;有42 766个,占86.8%的村拥有机电井;有4420个,占9.0%的村落有能够使用的灌溉用水塘和水库;有3835个,占7.8%的村具有排灌站。同时,全省农村年末实有机井数量为94.39万眼,与1995年相比,净增14.59万眼,平均每公顷拥有有效灌溉农田实有机井约0.21眼。

从灌溉水源的情况来看(见图2-5),全省有38 853个,占78.89%的村灌溉主要水源是地下水;只有6972个,占14.16%的村灌溉主要水源是地表水;不可忽视的是全省目前尚有3423个,占6.95%的村无

表 2-2　全省有农田水利设施的村个数及比重

水利设施项目	村个数/个	比重/%
在正常年景下用水有保障	42 942	87.2
有机电井	42 766	86.8
有能够使用的灌溉用水塘和水库	4420	9.0
有排灌站	3835	7.8

灌溉水源。同时，普查数据还显示，河北省农村在水利建设过程中，有 35 241 个，占 71.6%的村无资金来源；有 7850 个村，仅占 15.9%的村水利建设资金来源于集体；只有 2910 个，占 5.9%的村资金来源于国家；另外也有 3247 个村，占 6.6%的村水利建设资金来源于其他村。

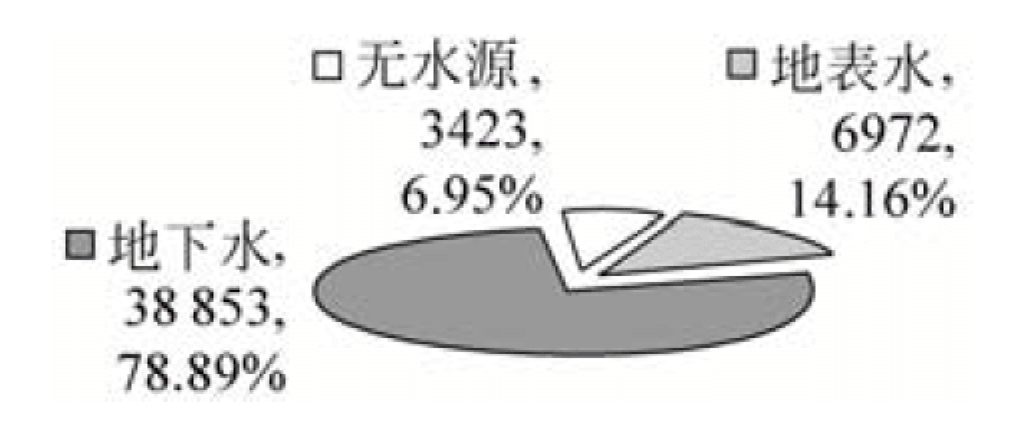


图 2-5　按灌溉主要水源分的村个数及比重图

3. 大牲畜养殖数量呈递减趋势

畜力在农村作为一种非常原始又非常实用的劳力工具，其实质是一种常规的可再生能源。普查数据显示，2006 年，全省农村养殖大牲畜 613.00 万头，牛 458.93 万头，马 28.99 万头，驴 90.16 万头，骡 34.92 万头。再结合《河北经济年鉴》数据（见图 2-6）可以看到，大牲畜、牛、马、驴、骡的数量均呈现递减态势，大牲畜数量由 1995 年的 870.88 万头减少至 2006 年的 613.00 万头，2012 年大牲畜数量为 498.15 万头，仅为 1995 年的 57.2%。其中牛由 1995 年的 579.34 万头，逐年递减至 2006 年的 458.93 万头，到 2012 年为 403.10 万头，是 1995 年的 69.6%；马由 1995 年的 48.68 万头，逐年递减至 2006 年的 28.99 万头，到 2012 年为 18.44 万头，是 1995 年的 37.9%；驴的数量从 1995 年的 167.69 万头，开始逐年减少到 2006 年的 90.16 万头，到 2012 年为 55.49 万头，是 1995 年的 33.1%；骡的数量也从 1995 年的 75.17 万头，逐年减至 2006 年的 34.92 万头，到 2012 年为 21.12 万头，是 1995 年

的 28.1%。

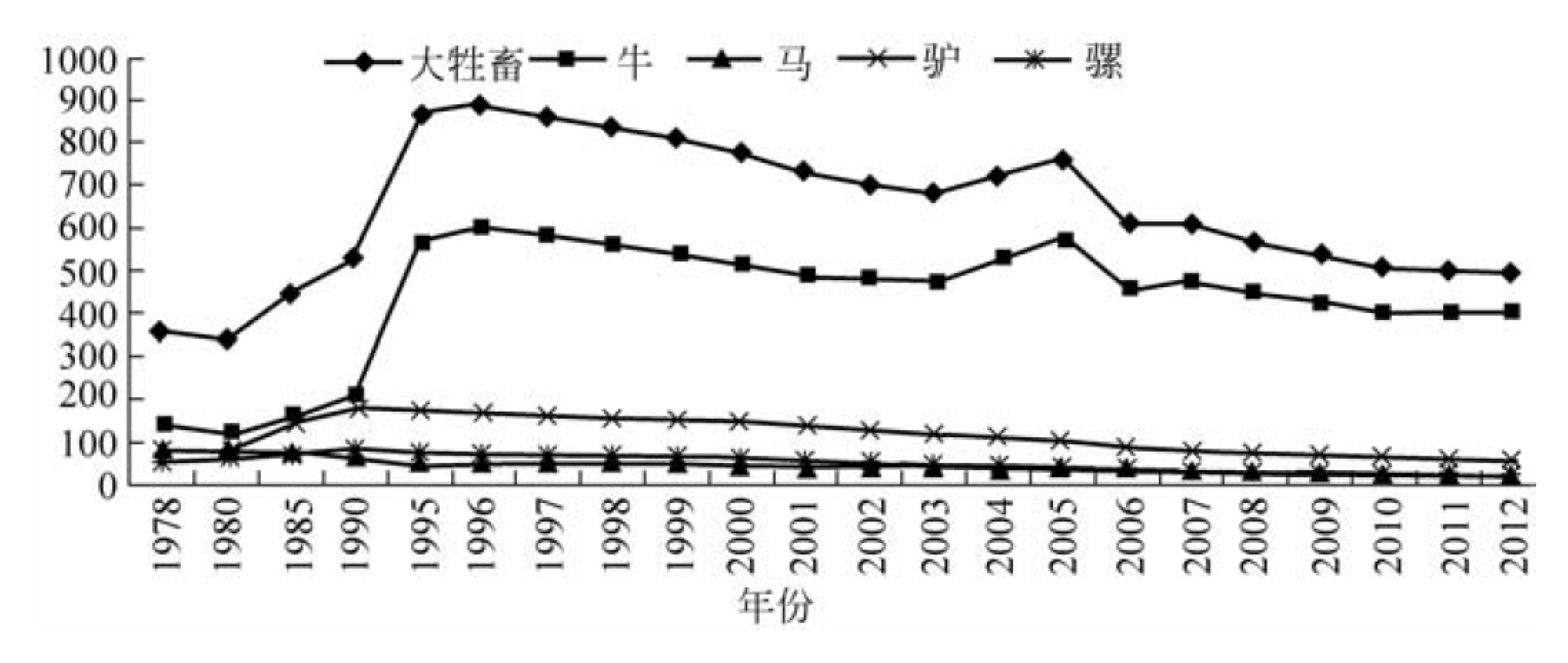


图 2-6　农村大牲畜数量时序图

4. 机械化生产日益成为全省农业生产的主要工具

随着科技的不断发展，农民收入的不断增加，高效率生产的农用机械化工具日益替代原始化生产的大牲畜成为农村生产的主要工具，这不仅使农民的生产效率大幅提高，也增加了农民对能源的消费需求，同时，这也是造成全省农村大牲畜养殖呈递减趋势的重要原因。

经济数据显示（见表 2-3），2012 年河北省农业机械总动力为 10 553.8 万 kW，与 2000 年的 7000.4 万 kW 相比，净增 50.76 个百分点。特别突出的是，农村拥有大中型拖拉机由 2000 年的 6.4 万台，增加到 2012 年的 21.4 万台，净增 234.38 个百分点；拥有大中型拖拉机配套农具由 2000 年的 10.8 万台增加到 2012 年的 41.0 万台，净增 279.63 个百分点；拥有联合收割机由 2000 年的 4.2 万台增加到 2012 年的 10.1 万台，净增 140.48 个百分点。农用机械化工具的普及使河北省机械化耕作面积也日益扩大，2000 年全省有 50.43% 和 29.79% 的播种面积使用机播和机收，到 2012 年，机播面积占播种面积的比重上升为 75.03%，机收面积占比上升为 47.91%。

表 2-3　全省农业机械数量表

指　标	2000 年	2005 年	2010 年	2011 年	2012 年
农业机械总动力/万 kW	7000.4	8487.2	10 151.3	10 349.2	10 553.8
大中型拖拉机/万台	6.4	10.1	17.3	19.8	21.4
小型拖拉机/万台	129.8	144.7	150.5	149.1	146.3
大中型拖拉机配套农具/万台	10.8	18.4	34.5	38.0	41.0
小型拖拉机配套农具/万台	156.7	191.6	201.0	199.4	195.5
节水灌溉机械/万套	3.1	3.5	4.4	4.5	4.9
联合收割机/万台	4.2	5.6	7.9	8.6	10.1
农用运输汽车/万辆	160.4	252.2	268.2	268.5	269.8

综上，现代机械化生产的日益普及在一定程度上降低了对畜力的依赖，降低了农民的劳动强度，提高了农业生产的效率，增加了农民的收入，有效地加快了全省新农村建设的步伐。但这种现代化农业生产却因水利建设资金的匮乏而主要以消耗地下水源作为代价，这与河北省水资源尤其是地下水资源缺乏的省情是极不相符的。再加上全省乡、村及村以下办水电站装机容量较全省农村电量需求杯水车薪的状况，严重阻碍了河北省新农村建设的发展。

第3章 河北省农民能源使用现状分析

3.1 指标选取与研究方法的选择

3.1.1 指标选取

指标体系的设置一般从两个方面入手：一是建立基础、详细的统计指标体系，使之成为了解农民能源使用情况的基础数据库；二是建立综合性的指标体系，进行综合评价、比较和分析。河北省第二次农业普查从生产、生活、农村能源基础设施建设等方面较科学、全面、详细地对全省各地区农民能源使用状况进行了摸底。因此本研究指标体系的选取完全来自河北省第二次农业普查中现有的指标数据。

但是，普查过程中，有部分指标数据是从各地区某种能源的使用农户总量来反映农民能源使用状况的，例如，农户年末农业机械拥有量，各地区按炊事用能源分的农村常住户户数，各地区按饮用水源分的农村常住户户数等。因为地区间农村常住人口户数是存在差异的，比如保定市拥有220.97万户农村常住户，而秦皇岛市仅拥有农村常住户58.47万户，如果仅按照能源使用的绝对户数来进行分析评价，结果将因地区农户数量差异而产生较大误差，因此在分析过程中我们将普查所得的绝对指标数据统一转换成每百户拥有量，如每百户年末农业机械拥有量、每百户无取暖设施户数、每百户主要炊事能源为柴草户数，这样可以真实地反映各地区农民能源的使用状况。

我们在分析中从两个方面进行综合评价，第一个方面是分析比较全省各地区农民生活能源的使用情况，以找出河北省各地区农民生活能源使用差异；第二方面是分析比较全省各地区农村能源基础设施建设情况及农民生产用能状况。因此在分析过程中将分别从这两个方面

对河北省 11 个地区的农民能源使用状况进行比较分析。

3.1.2　研究方法的选择

本书中在测算农民能源使用状况时使用的是多元统计分析方法中的因子分析方法。因子分析的概念起源于 20 世纪初 Kan Pearson 和 Charles Spearmen 等人关于智力测验的统计分析。近年来，随着现代高速电子计算机的出现，人们将因子分析的理论成功地应用于心理学、医学、气象、地质、经济学等领域，使得因子分析的理论和方法更加丰富。

在各个领域的科学研究中往往需要对反映事物的多个变量进行大量的观测，收集大量数据以便进行分析寻找规律。多变量大样本无疑会为科学研究提供丰富的信息，但也在一定程度上增加了数据采集的工作量，更重要的是在大多数情况下，许多变量之间可能存在相关性而增加了问题分析的复杂性。由于各变量间存在一定的相关关系，因此有可能用较少的综合指标分别综合存在于各变量中的各类信息，而综合指标之间彼此不相关，即各指标代表的信息不重叠。因子分析(factor analysis)就是从研究相关矩阵内部的依赖关系出发，把一些具有错综复杂关系的变量归纳为少数几个综合因子的一种多变量统计分析方法。代表各类信息的综合指标就称为因子或主成分。因子分析的基本思想是根据相关性的大小把变量分组，使得同组内的变量之间相关性较高，而不同组的变量相关性较低，每组变量代表一个基本结构，这个基本结构称为公共因子。对于所研究的问题就可以试图用最少个数的不可测的公共因子的线性函数与特殊因子之和来描述原来观测的每一个分量。也就是从一些错综复杂关系的现象中找出少数几个主要因子，每一个主要因子就代表反映经济变量之间相互依赖的一种经济作用，抓住这些主要因子就可以帮助我们对复杂的经济问题进行分析和解释。因子分析方法的特点是在尽可能多地考虑各种评价指标的前提下，通过指标贡献率的计算和数据处理后，可以只用少数几个新的综合指标来代表原来所有指标所包含的信息，同时也能客观确定各个指标的权数，避免主观随意性。

1. 初始因子模型

因子分析分为两类，即 R 型因子分析（对变量作因子分析）和 Q 型因子分析（对样本作因子分析）。因为本研究主要是探讨地区间农民能源使用状况的差异，即为变量的因子分析，因此这里介绍 R 型因子分析的数学模型。设有 n 个样本，每个样本观测 p 个变量 $x_1, x_2, \cdots, x_p$，为了对变量进行比较并消除由于观测量量纲的差异及数量级所造成的影响，将样本观测数据进行标准化处理，使标准化后的变量的均值为 0，方差为 1，其标准化指标为

$$X_i = \frac{x_i - \bar{x}_i}{s_i}, \quad i = 1, 2, \cdots, p \tag{3-1}$$

其中，$\bar{x}_i$ 为所有变量的平均数，s_i 为标准差。设原公共因子变量为 y_1，$y_2, \cdots, y_m$，经标准化后的公共因子变量记为 $F_1, F_2, \cdots, F_m (m < p)$。如果：

（1）$\boldsymbol{X} = (X_1, X_2, \cdots, X_p)'$ 是可观测的随机变量，且均值向量 $E(\boldsymbol{X}) = \boldsymbol{0}$，协方差阵 $\mathrm{cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$，且协方差阵 $\boldsymbol{\Sigma}$ 与相关矩阵 $\boldsymbol{R}$ 相等；

（2）$\boldsymbol{F} = (F_1, F_2, \cdots, F_m)' (m < p)$ 是不可测的向量，其均值向量 $E(\boldsymbol{F}) = \boldsymbol{0}$，协方差阵 $\mathrm{cov}(\boldsymbol{F}) = \boldsymbol{I}$；

（3）$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p)'$，$\boldsymbol{\varepsilon}$ 与 $\boldsymbol{F}$ 相互独立，且 $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$，其协方差阵 $\boldsymbol{\Sigma}_{\varepsilon}$ 是对角阵，则

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \varepsilon_2 \\ \qquad \vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + \varepsilon_p \end{cases} \tag{3-2}$$

称此模型为因子模型，称 $F_1, F_2, \cdots, F_m$ 为公共因子，$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$ 为特殊因子，a_{ij} 为因子载荷。

设

$$x_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + \varepsilon_i, \quad i = 1, 2, \cdots, p$$

则有 $E(x_iF_j) = \sum_{k=1}^{m} a_{ik}E(F_kF_j) = \sum_{k=1}^{m} a_{ik}r_{(F_kF_j)} = a_{ij}$，由于 F_k、F_j 不相关，且 $r_{(F_jF_1)} = 1$，即 $a_{ij} = r_{x_i, F_j}$，因子载荷 a_{ij} 是第 i 个变量与第 j 个公

共因子的相关系数。因子载荷矩阵中列的平方和

$$s_j = \sum_{i=1}^{p} a_{ij}, \quad j = 1,2,\cdots,p \tag{3-3}$$

称 s_j 为公共因子 F_j 对 x_i 的贡献，是衡量公共因子相对重要性的指标。

通过因子分析，取累计贡献率达到一定值以上的所有特征根，确定公共因子，即提取几个新的综合变量，使得这些较少的变量既尽可能地反映原来变量的统计特性，又在新变量之间保持相互的独立性。

2．因子旋转

建立因子分析模型的目的不仅是找出主因子，更重要的是知道每个因子的意义，以便对实际问题进行分析。但通过初始因子模型而求得的主因子解，各主因子的典型代表变量不很突出，容易使因子的意义含糊不清，不便于对实际问题进行分析。由线性代数知识知道，一个正交变换对应坐标系的旋转，而且主因子的任一解都可由初始因子模型求得的主因子解经过旋转（右乘一个正交矩阵）而得出。因子旋转有方差最大正交旋转和斜交旋转，比较常用的是方差最大化正交旋转。经过旋转后，公共因子对 X_i 的贡献率不改变，但公共因子本身可能有较大变化，从而可通过适当的旋转得到比较满意的公共因子，实现这一目的的变换方法称为因子旋转。设从公因子 F 旋转到公因子 G，则有

$$X_i = \sum_{j=1}^{m} b_{ij} G_{ij} + \varepsilon_i, \quad i = 1,2,\cdots,p; j = 1,2,\cdots,m; m < p \tag{3-4}$$

3．因子得分模型

因子模型建立以后，反过来考察每个样本，即将指标表示为公共因子的线性组合变换为将公共因子表示为指标的线性组合，这样就可以从指标的观测值估计各公共因子的值，这组值称为因子得分。

设公共因子 F 由变量 X 表示的线性组合为

$$F_j = \beta_{j1} X_1 + \beta_{j2} X_2 + \cdots + \beta_{jp} X_p, \quad j = 1,2,\cdots,m \tag{3-5}$$

根据每个观测量在公共因子上的得分以及每个公共因子所对应的贡献率计算出观测量的最终得分。然后对计算得分结果进行排序，并

确定各地区的农民能源使用状况。

3.2 各地区农民生活能源使用状况实证研究

因为河北省第二次农业普查涉及的全省农民生活能源状况的指标多达29个，反映问题较全面，指标变量间很可能存在强的相关性。但又不能进行主观判断，随便增减指标。尤其是在综合评分分析确定权重时，如果权重不准确，会造成分析评价结果出现比较大的偏差，这种误差是十分严重的。为了避免此种情况的发生，采用因子分析方法进行分析评价，因子分析方法除了能有效避免主观确定权重的误差以及之前所述的优点，它还能根据各个指标的具体情况对每个指标进行自动归类，这样就避免了因主观判断造成指标错误归类而影响综合比较分析的结果。

1. 原始指标数据标准化处理

在对指标体系进行因子分析之前，必须对各个指标进行标准化无量纲处理，此处采用的标准化处理公式为

$$X_i = \frac{x_i - \bar{x}_i}{s_i}, \quad i = 1,2,\cdots,p \tag{3-6}$$

其中，$\bar{x}_i$ 为所有变量的平均数，s_i 为标准差，x_i 为标准化以前的变量值，X_i 为标准化之后的变量值。标准化以后的变量的均值为0，标准差为1。标准化以后的指标数据见表3-1。

2. 关于河北省各地区农民生活能源使用状况因子分析

本部分的数据处理利用软件SPSS 19.0进行。确定因子个数时，原则上选取主成分特征值大于1，并且所选取公共因子的累计方差贡献率达85%以上。计算结果显示，相关矩阵的前7个特征值大于1的因子其累计方差贡献率达95.837%。此时，研究考虑到后面将要阐述的因子载荷矩阵和公共因子的命名等问题，仅选取相关矩阵特征值大于1的前5个因子作为公共因子，如表3-2所示。

表 3-1　河北省各地区农民生活能源使用状况指标数据标准化表

指标(百户数)	石家庄	唐　山	秦皇岛	邯　郸	邢　台	保　定	张家口	承　德	沧　州	廊　坊	衡　水
X_1：主要炊事用柴草	−1.4242	0.5933	0.7352	−1.6100	−1.0803	−0.4846	−0.0433	1.1184	1.0833	0.4137	0.6984
X_2：主要炊事用煤	1.3534	−0.8127	−0.5795	1.6280	1.1520	0.4555	0.1548	−1.0113	−0.9994	−0.8165	−0.5243
X_3：主要炊事用煤气、天然气	−0.3608	1.3449	−0.4953	−0.7723	−0.8161	−0.0079	−0.6256	−0.1008	0.0625	2.3988	−0.6274
X_4：主要炊事用沼气	2.1805	0.9852	−0.7120	1.0860	0.0116	−0.1622	−0.8561	−0.8939	−0.6238	−0.7581	−0.2573
X_5：主要炊事用电	1.0839	0.4826	−0.5100	−1.1666	−0.1766	−0.6760	1.6471	1.2791	−0.8958	−1.1509	0.0832
X_6：主要炊事用太阳能	2.2250	−0.9482	−1.2368	0.7766	0.2698	0.4180	0.0326	−0.8627	−0.2018	−0.8965	0.4240
X_7：辅助炊事用柴草	0.9056	−0.7111	−0.7335	1.2952	1.4978	1.1317	−0.1956	−1.2088	−0.9446	−0.4658	−0.5708
X_8：辅助炊事用煤	−0.7582	−0.6663	−0.0028	−1.0780	−0.0583	0.8918	1.1049	−1.1005	−0.7819	0.5253	1.9240
X_9：辅助炊事用煤气、天然气	−1.1179	1.4497	0.0957	−0.7013	−0.9337	−0.4659	−1.1093	0.5691	1.6818	0.7829	−0.2510
X_{10}：辅助炊事用沼气	1.1405	1.7139	0.4335	1.2203	−0.1016	−0.8124	−1.0637	0.0744	−0.8627	−0.9588	−0.7835
X_{11}：辅助炊事用电	−0.2775	0.2510	0.7174	−0.8915	−0.6286	−0.5110	1.7353	1.8066	−0.7578	−0.7599	−0.6839
X_{12}：辅助炊事用太阳能	2.4096	−0.8016	−1.0522	−0.5597	1.0602	−0.1742	−0.2870	−0.4765	0.5946	−0.5289	−0.1841
X_{13}：彩电	0.3251	0.7565	−0.0008	−0.1834	−0.2479	0.4951	−1.9488	−1.2505	0.6151	1.7699	−0.3303
X_{14}：汽车	−0.0818	0.4263	−0.1579	−0.6700	−0.5100	0.0222	−0.9341	−0.5300	0.4183	2.6989	−0.6820
X_{15}：摩托车	−0.1283	0.9265	−1.1234	−0.0448	1.0934	−1.3527	−0.5894	−1.2559	0.4508	0.4375	1.5864
X_{16}：手机	0.6330	1.3138	0.2482	−0.7108	−0.7876	1.0017	−1.6544	−0.6655	−0.0591	1.4052	−0.7245

续表

指标(百户数)	石家庄	唐　山	秦皇岛	邯　郸	邢　台	保　定	张家口	承　德	沧　州	廊　坊	衡　水
X_{17}：计算机	1.2265	0.9314	−0.3108	−0.7290	−0.5076	−0.5445	−1.0242	−0.7044	0.1442	2.1244	−0.6060
X_{18}：无取暖户数	−0.0048	−0.6374	−0.6418	2.3400	1.3065	−0.5110	−0.6249	−0.6211	−0.5522	−0.5992	0.5458
X_{19}：空调取暖户数	0.6771	−0.5186	−1.1191	0.5234	0.8522	−0.4010	−1.2924	−1.1820	0.5980	1.8532	0.0091
X_{20}：暖气取暖户数	1.4819	−0.3179	−1.2265	−1.1020	0.1845	0.7790	−1.2470	−0.9057	0.7086	0.5852	1.0599
X_{21}：火炕取暖户数	−1.0008	0.8659	1.5460	−1.3508	−1.2277	−0.0316	0.6130	1.2633	0.0086	0.0561	−0.7422
X_{22}：使用入户管道水	0.4452	−0.4371	−1.5987	−0.1039	0.6717	−0.6226	−0.2606	−1.5038	1.1304	1.1531	1.1263
X_{23}：获取饮用水困难	−0.4862	−1.1652	0.1831	0.3391	−0.4244	−0.2880	1.7263	0.9304	1.3397	−1.2042	−0.9507
X_{24}：饮用水为净化水	0.3598	0.1092	−0.2512	0.0158	−0.4100	−0.4586	0.4060	−0.7128	−0.4752	2.6468	−1.2299
X_{25}：饮用水为深井水	−0.1028	0.4252	−0.2523	−0.3545	0.5206	0.8681	−1.6044	−0.8535	0.9863	−1.2744	1.6418
X_{26}：饮用水为浅井水	−0.2635	−0.5710	0.7389	0.0834	−0.2065	−0.6284	1.3514	2.0637	−0.6327	−1.3253	−0.6099
X_{27}：饮用水为江河湖水	−0.3482	−0.6339	−0.5254	1.8111	−0.5835	0.5088	1.9885	−0.2327	−0.6638	−0.6474	−0.6735
X_{28}：饮用水为池塘水	0.1850	−0.8381	−0.4644	1.5051	2.0946	−0.3677	0.5291	−0.2466	−0.7600	−0.8470	−0.7901
X_{29}：饮用水为雨水	0.1349	−0.4428	−0.3855	2.9676	−0.1706	−0.3363	−0.2327	−0.1988	−0.4441	−0.4465	−0.4453

表 3-2　主成分方差贡献率

因子	旋转前			旋转后		
	特征根	贡献率/%	累计贡献率/%	特征根	贡献率/%	累计贡献率/%
1	8.683	29.941	29.941	8.514	29.360	29.360
2	8.573	29.561	59.502	8.293	28.598	57.958
3	3.333	11.494	70.996	3.227	11.129	69.087
4	2.440	8.415	79.411	2.816	9.711	78.798
5	2.230	7.689	87.100	2.408	8.302	87.100

3. 计算影响力度

由于各指标变量之间相关性非常强，为了使各变量之间经济意义的区别更为清晰，因此在分析过程中以“四次方差极大化”为准则进行因子正交旋转，从而得到方差极大化后的因子载荷矩阵。可以发现，旋转后的总体因子的累计方差贡献率仍为 87.100%。根据旋转后的因子载荷矩阵可以得到具有明确经济意义的公共因子，如表 3-3 所示。

表 3-3　旋转后的因子载荷矩阵

指标(百户数)	公共因子				
	F_1	F_2	F_3	F_4	F_5
X_1：主要炊事用柴草	−0.1045	−0.9431	−0.1424	−0.1901	−0.0999
X_2：主要炊事用煤	0.0010	0.9673	0.0212	0.1491	0.0995
X_3：主要炊事用煤气、天然气	0.5152	−0.5416	0.5901	0.1039	−0.0795
X_4：主要炊事用沼气	0.3091	0.5446	−0.0574	0.6428	0.3426
X_5：主要炊事用电	−0.6118	−0.0598	−0.0010	0.0400	0.6766
X_6：主要炊事用太阳能	0.1706	0.7759	−0.1404	0.0152	0.5024
X_7：辅助炊事用柴草	0.1733	0.8892	0.0225	0.0599	0.0377
X_8：辅助炊事用煤	0.0763	−0.0806	−0.0852	−0.7702	0.1050
X_9：辅助炊事用煤气、天然气	0.3053	−0.7953	−0.0337	0.2055	−0.2719
X_{10}：辅助炊事用沼气	−0.0030	0.2035	−0.0715	0.9370	0.0076
X_{11}：辅助炊事用电	−0.8781	−0.3522	0.1735	0.0503	0.2060
X_{12}：辅助炊事用太阳能	0.2792	0.4795	−0.0906	0.0079	0.7162

续表

指标（百户数）	公共因子				
	F_1	F_2	F_3	F_4	F_5
X_{13}：彩电	0.8863	−0.2126	0.2390	0.2274	−0.1085
X_{14}：汽车	0.6578	−0.4034	0.6185	−0.0185	−0.1028
X_{15}：摩托车	0.6339	0.0875	−0.3228	−0.1743	−0.0192
X_{16}：手机	0.6375	−0.3788	0.3275	0.4275	0.1066
X_{17}：计算机	0.6907	−0.2545	0.5463	0.2545	0.2466
X_{18}：无取暖户数	0.1393	0.8159	−0.2824	0.0803	−0.3908
X_{19}：空调取暖户数	0.8437	0.3173	0.2559	−0.1127	−0.0702
X_{20}：暖气取暖户数	0.7592	0.0507	−0.1639	−0.1885	0.5406
X_{21}：火炕取暖户数	−0.4951	−0.8181	0.1829	0.1382	−0.0445
X_{22}：使用入户管道水	0.7351	0.2325	−0.0337	−0.4693	0.1094
X_{23}：获取饮用水困难	−0.7246	0.0403	−0.0198	−0.2247	−0.0908
X_{24}：饮用水为净化水	0.3340	−0.0337	0.9248	−0.0431	−0.0581
X_{25}：饮用水为深井水	0.4825	−0.0321	−0.8393	−0.0126	0.0616
X_{26}：饮用水为浅井水	−0.9582	−0.0439	−0.0156	0.0829	0.0368
X_{27}：饮用水为江河湖水	−0.5587	0.5533	0.2530	−0.1281	−0.2584
X_{28}：饮用水为池塘水	−0.2362	0.8452	−0.0265	0.0145	−0.1485
X_{29}：饮用水为雨水	−0.0935	0.7108	0.0408	0.3355	−0.4976

根据每个指标在5个因子上的载荷情况，描述河北省农民生活能源使用状况的29个指标一共被分成5大类，结果如下：

第一类：X_5—每百户主要炊事用电户数；X_{11}—辅助炊事用电户数比重；X_{13}—每百户彩电台数；X_{14}—每百户汽车辆数；X_{15}—每百户摩托车辆数；X_{16}—每百户手机部数；X_{17}—每百户计算机台数；X_{19}—每百户空调取暖户数；X_{20}—每百户暖气取暖户数；X_{22}—使用入户管道水户数比重；X_{23}—获取饮用水困难比重；X_{26}—每百户饮用水为浅井水户数。这个公共因子主要反映农民在耐用消费品等较高质量生活方面的用能情况，本研究取为耐用消费品用能。这些耐用消费品一般能源利用率相对较高，浪费较少，污染较少。因此通过这个公共因子不仅可以反映河北省各个地区农民生活质量和生活水平的状况，也可以反映农民能

源使用的效率。

第二类：X_1—每百户主要炊事用柴草户数；X_2—每百户主要炊事用煤户数；X_3—每百户主要炊事用煤气、天然气户数；X_6—每百户主要炊事用太阳能户数；X_7—每百户辅助炊事用柴草户数；X_9—每百户辅助炊事用煤气、天然气户数；X_{18}—每百户无取暖户数；X_{21}—每百户火炕取暖户数；X_{27}—每百户饮用水为江河湖水户数；X_{28}—每百户饮用水为池塘水户数；X_{29}—每百户饮用水为雨水户数。这个公共因子主要反映农民对能耗较高、健康程度低、污染严重能源的使用状况，本研究取名为待改善能源。因此这个公共因子在很大程度上与经济、社会、环境的发展是呈负相关的，即这个因子得分越高，地区农民生活能源的使用越亟需改善，为一个负相关因子。

第三类：X_{24}—每百户饮用水为净化水户数；X_{25}—每百户饮用水为深井水户数。这个公共因子主要反映对相对健康饮用水源的使用状况，本研究取名为健康水资源。

第四类：X_4—每百户主要炊事用沼气户数；X_8—每百户辅助炊事用煤户数。X_{10}—每百户辅助炊事用沼气户数。这个公共因子主要反映农民对沼气的使用状况，本研究取名为沼气能。

第五类：X_6—每百户主要炊事用太阳能户数；X_{12}—每百户辅助炊事用太阳能户数。这个公共因子主要反映农民对太阳能的使用状况，本研究取名为太阳能。通过第四类和第五类公共因子可以发现河北省 11 个地区农村新能源建设的状况。

4. 因子得分计算

由式(3-5)可知，河北省各地区农民生活能源使用状况的因子得分函数可表示为

$$F_i = \sum_{j=1}^{29} \beta_{ij} x_j, \quad i = 1,2,\cdots,5 \tag{3-7}$$

其中 F_i 表示第 i 个因子得分，β_{ij} 表示第 j 个指标对第 i 个因子的贡献程度，x_j 表示第 j 个指标。

因此，结合表 3-4 可以得出 5 个公共因子的得分函数为

$$\begin{cases} F_1 = -0.0002X_1 - 0.0083X_2 + 0.0469X_3 + \cdots - 0.0052X_{29} \\ F_2 = -0.1220X_1 + 0.1208X_2 - 0.0454X_3 + \cdots + 0.0922X_{29} \\ F_3 = -0.0850X_1 + 0.0501X_2 + 0.1512X_3 + \cdots + 0.0165X_{29} \\ F_4 = -0.0366X_1 + 0.0278X_2 + 0.0185X_3 + \cdots + 0.1000X_{29} \\ F_5 = -0.0327X_1 + 0.0303X_2 - 0.0180X_3 + \cdots - 0.2164X_{29} \end{cases} \tag{3-8}$$

表 3-4　旋转后的因子得分系数矩阵

指标(百户数)	F_1 耐用消费品用能	F_2 待改善能源	F_3 健康水资源	F_4 沼气能	F_5 太阳能
X_1：主要炊事用柴草	−0.0002	−0.1220	−0.0850	−0.0366	−0.0327
X_2：主要炊事用煤	−0.0083	0.1208	0.0501	0.0278	0.0303
X_3：主要炊事用煤气、天然气	0.0469	−0.0454	0.1512	0.0185	−0.0180
X_4：主要炊事用沼气	0.0264	0.0440	−0.0269	0.2250	0.1315
X_5：主要炊事用电	−0.0912	−0.0153	0.0390	0.0215	0.3040
X_6：主要炊事用太阳能	0.0095	0.0853	0.0004	−0.0062	0.1941
X_7：辅助炊事用柴草	0.0150	0.1132	0.0435	−0.0038	0.0005
X_8：辅助炊事用煤	0.0128	0.0044	0.0097	−0.2761	0.0376
X_9：辅助炊事用煤气、天然气	0.0482	−0.1078	−0.0818	0.0966	−0.1125
X_{10}：辅助炊事用沼气	−0.0024	−0.0027	−0.0655	0.3428	0.0020
X_{11}：辅助炊事用电	−0.1167	−0.0370	0.0771	0.0226	0.1204
X_{12}：辅助炊事用太阳能	0.0172	0.0461	0.0072	−0.0030	0.2882
X_{13}：彩电	0.1029	−0.0247	0.0256	0.0727	−0.0557
X_{14}：汽车	0.0637	−0.0231	0.1688	−0.0324	−0.0329
X_{15}：摩托车	0.0891	−0.0026	−0.1200	−0.0511	−0.0370
X_{16}：手机	0.0643	−0.0491	0.0531	0.1477	0.0482
X_{17}：计算机	0.0599	−0.0198	0.1473	0.0692	0.1116
X_{18}：无取暖户数	0.0321	0.0943	−0.0803	0.0207	−0.1905
X_{19}：空调取暖户数	0.0949	0.0536	0.0788	−0.0673	−0.0466
X_{20}：暖气取暖户数	0.0854	−0.0075	−0.0510	−0.0621	0.2042
X_{21}：火炕取暖户数	−0.0622	−0.0976	0.0309	0.0646	0.0115
X_{22}：使用入户管道水	0.0888	0.0361	0.0041	−0.1791	0.0221

续表

指标(百户数)	F_1 耐用消费品用能	F_2 待改善能源	F_3 健康水资源	F_4 沼气能	F_5 太阳能
X_{23}：获取饮用水困难	−0.0850	0.0151	0.0290	−0.0804	−0.0230
X_{24}：饮用水为净化水	0.0109	0.0413	0.3067	−0.0669	−0.0004
X_{25}：饮用水为深井水	0.0863	−0.0482	−0.3052	0.0407	−0.0145
X_{26}：饮用水为浅井水	−0.1175	−0.0046	0.0220	0.0361	0.0397
X_{27}：饮用水为江河湖水	−0.0727	0.0937	0.1316	−0.0750	−0.0956
X_{28}：饮用水为池塘水	−0.0284	0.1109	0.0344	−0.0159	−0.0699
X_{29}：饮用水为雨水	−0.0052	0.0922	0.0165	0.1000	−0.2164

由得分函数可以算出反映河北省农民生活能源使用状况的 5 个公共因子在全省 11 个样本地区上的因子得分，得分结果及各公共因子排名见表 3-5。

这 5 个公共因子的旋转方差贡献率分别为：29.360%、28.598%、11.129%、9.711%、8.302%，则它们在反映全省各地区农民生活能源使用状况综合评价时的权重公式可表示为

$$\omega_i = \lambda_i / \sum_{i=1}^{5} \lambda_i \tag{3-9}$$

其中 ω_i 表示公共因子的权重，λ_i 表示各因子的方差贡献率。我们可以得到 5 个公共因子的权重分别为 0.337 08、0.328 34、0.127 77、0.111 49、0.095 32。因此，河北省各地区农民生活能源使用状况综合评价总得分为

$$F_{\text{total}} = \sum_{i=1}^{5} \omega_i F_i \tag{3-10}$$

其中 F_{total} 表示地区综合得分，ω_i 表示公共因子的权重，F_i 表示各公共因子的得分。各地区综合得分见表 3-5。

表 3-5　公共因子得分情况[①]

地　区	F_1 耐用消费品用能	排名	F_2 待改善能源	排名	F_3 健康水资源	排名	F_4 沼气能	排名	F_5 太阳能	排名	综合得分	综合排名
石家庄	0.483	4	1.052	3	0.421	3	1.026	2	2.323	1	0.207	4
唐山	0.461	5	−1.091	11	−0.174	8	1.668	1	0.009	6	0.6781	2
秦皇岛	−0.768	9	−0.890	9	−0.115	7	0.594	4	−0.623	10	0.0256	5
邯郸	−0.153	8	1.868	1	0.051	4	0.895	3	−1.887	11	−0.7385	11
邢台	0.357	6	1.166	2	−0.606	9	−0.460	7	0.004	7	−0.3908	9
保定	0.188	7	0.120	5	−0.111	6	−0.316	6	0.147	5	−0.0115	6
张家口	−1.839	11	0.342	4	0.923	2	−1.421	11	0.455	2	−0.7294	10
承德	−1.528	10	−0.929	10	−0.068	5	0.523	5	0.161	4	−0.1448	8
沧州	0.679	2	−0.763	8	−0.767	10	−0.487	8	−0.263	8	0.3023	3
廊坊	1.482	1	−0.660	7	2.205	1	−0.761	9	−0.535	9	0.862	1
衡水	0.637	3	−0.215	6	−1.758	11	−1.262	10	0.208	3	−0.06	7

① 说明：公共因子 F_2 待改善能源因为是负相关因子，因此在测算综合得分时将其取负处理。

表 3-6　河北省各地区农民生产能源使用状况指标数据标准化表

指　　标	石家庄	唐　山	秦皇岛	邯　郸	邢　台	保　定	张家口	承　德	沧　州	廊　坊	衡　水
y_1：通电的村	−0.577	−0.983	0.741	−1.003	0.538	0.376	−2.078	0.924	0.924	0.417	0.721
y_2：灌溉主要水源是地表水	0.925	−0.193	1.514	−0.704	0.201	−0.363	0.168	1.535	−0.494	−1.584	−1.005
y_3：灌溉主要水源是地下水	−0.212	0.447	−0.787	0.367	0.301	0.289	−1.837	−1.519	0.612	1.297	1.043
y_4：灌溉无水源的村	−0.441	−0.513	−0.062	0.027	−0.616	−0.136	2.620	1.021	−0.512	−0.647	−0.741
y_5：正常年景用水有保障	0.131	0.707	0.069	−0.054	0.580	0.163	−2.525	−1.125	0.477	0.804	0.772
y_6：有机电井的村	0.129	0.472	−0.957	0.378	0.621	0.139	−1.756	−1.689	0.899	0.911	0.852
y_7：有灌溉用水塘和水库村	0.661	−0.122	2.289	−0.574	0.216	−0.906	−0.627	−0.333	1.086	−0.840	−0.850
y_8：有排灌站的村	1.781	0.552	−0.306	−0.278	1.462	−0.659	0.611	−0.696	−0.092	−1.402	−0.974
y_9：来源于国家的村	−0.060	−0.453	−1.373	0.976	0.415	−0.314	0.688	2.108	−0.199	−0.640	−1.148
y_{10}：来源于集体的村	0.241	1.842	−1.232	−0.180	0.039	0.552	−0.607	−1.455	−0.507	1.394	−0.089
y_{11}：来源于其他的村	0.127	−0.853	−1.724	1.260	1.632	−0.212	−0.394	−0.562	−0.380	−0.063	1.169
y_{12}：无资金的村	−0.244	−1.180	2.266	−0.668	−0.802	−0.305	0.446	0.756	0.672	−1.003	0.063
y_{13}：乡、村办水电站个数	1.897	−0.207	−0.558	1.722	−0.821	0.582	−0.207	0.056	−0.821	−0.821	−0.821
y_{14}：乡、村办水电站发电量	2.970	−0.408	−0.419	−0.340	−0.419	−0.002	−0.132	0.009	−0.419	−0.419	−0.419
y_{15}：大中型拖拉机	1.571	−0.343	−1.434	0.451	0.653	−0.219	−1.766	−0.589	0.524	0.202	0.950
y_{16}：小型拖拉机	0.218	−0.535	−0.605	−0.532	0.594	−0.324	−0.819	−1.013	1.036	−0.395	2.374
y_{17}：节水灌溉机械	−0.826	0.880	2.254	−0.198	0.110	−1.010	−0.881	−1.107	0.530	−0.075	0.324
y_{18}：农用运输车辆	0.479	−0.156	−0.236	1.062	−0.385	1.292	−1.330	−1.630	0.596	1.134	−0.825
y_{19}：渔业机动船	−0.463	0.787	2.778	−0.451	−0.472	−0.491	−0.515	−0.323	0.052	−0.404	−0.497
y_{20}：联合收割机	1.532	−0.765	−1.189	1.019	0.653	0.393	−1.236	−1.318	0.179	−0.121	0.851
y_{21}：农村用电量	0.351	2.284	−0.781	−0.111	−0.587	−0.765	−1.151	−0.499	0.228	1.205	−0.174
y_{22}：役畜	−0.009	−0.050	−0.515	−0.672	−0.226	−1.666	0.825	0.691	0.455	−0.901	2.068
y_{23}：牛	0.940	0.037	−0.913	−1.044	−0.460	−1.666	0.171	1.658	0.169	−0.071	1.178
y_{24}：马	−0.201	−0.441	−1.011	−0.240	−0.345	−0.944	0.048	2.618	−0.065	−0.281	0.860
y_{25}：驴	−0.695	0.556	−0.765	−0.912	0.139	−1.454	1.242	−0.681	−0.114	1.195	1.488
y_{26}：骡	−0.317	−0.509	−0.865	0.116	−0.540	−1.015	2.554	0.980	−0.227	−0.141	−0.036

3.3 各地区农民生产能源使用状况实证研究

关于河北省农民生产能源使用状况的指标，主要是结合河北省第二次农业普查数据，从农村能源基础设施建设和农民生产用能两个方面考虑进行选取。各指标属于农村能源基础设施建设的是：通电村、灌溉主要水源是地表水、灌溉主要水源是地下水、正常年景用水有保障的村、有机电井的村、有灌溉用的水塘和水库的村、有排灌站的村、水利建设资金来源于国家的村、水利建设资金来源于集体的村、水利建设资金来源于其他的村、水利建设无资金来源的村、乡村水电站个数、乡村办水电站发电量。反映农民生产用的指标主要是：大中型拖拉机辆数、小型拖拉机辆数、节水灌溉机械套数、农用运输车辆数、渔业机动船艘数、联合收割机台数、农村用电量、役畜数以及牛、马、驴、骡的数量。

1. 指标数据标准化处理结果

由前面所介绍的数据标准化的式(3-6)，对河北省各地区农民生产能源使用状况指标数据的处理结果见表 3-6。

2. 河北省各地区农民生产能源使用状况的因子分析

确定公共因子个数时，与各地区农民生活能源使用状况的分析方法一样。因此，关于全省农民生产能源使用状况的因子分析仍选取相关矩阵特征值大于 1 的前 5 个因子作为公共因子，如表 3-7 所示。

表 3-7 主成分方差贡献率

因子	旋转前			旋转后		
	特征根	贡献率/%	累计贡献率/%	特征根	贡献率/%	累计贡献率/%
1	8.525	32.789	32.789	7.785	29.942	29.942
2	5.122	19.700	52.488	4.675	17.982	47.924
3	3.845	14.787	67.275	3.952	15.200	63.124
4	3.071	11.810	79.086	3.126	12.024	75.147
5	2.015	7.748	86.834	3.038	11.686	86.834

3. 计算影响力度

此处以“四次方差极大化”为准则进行因子正交旋转，从而得到方差极大化后的因子载荷矩阵。可以发现，旋转后的总体因子的累计方差贡献率仍为 86.834%。旋转后的公共因子如表 3-8 所示。

根据每个指标在 5 个公共因子上的载荷情况，将描述河北省农民生产能源使用状况的 26 个指标一共分成 5 大类，结果如下：

第一类：y_2—灌溉主要水源是地表水的村比重；y_3—灌溉主要水源是地下水的村比重；y_4—灌溉无水源的村比重；y_5—正常年景用水有保障的村比重；y_6—有机电井的村比重；y_{15}—每百户拥有大中型拖拉机辆数；y_{16}—每百户小型拖拉机拥有辆数；y_{20}—每百户联合收割机台数；y_{26}—每百户骡头数。这个公共因子主要反映各地区满足农业生产用能的基本状况，因此本研究取名为基本生产用能。

表 3-8　旋转后的因子载荷矩阵

指　　标	F_1 基本生产用能	F_2 水利建设	F_3 劳力用能	F_4 能源基础设施	F_5 能源保障
y_1：通电的村	0.503	0.230	0.125	−0.343	−0.627
y_2：灌溉主要水源是地表水	−0.503	0.434	0.094	0.484	−0.475
y_3：灌溉主要水源是地下水	0.909	−0.116	−0.123	−0.250	0.272
y_4：灌溉无水源的村	−0.952	−0.183	0.107	−0.022	−0.019
y_5：正常年景用水有保障	0.937	0.199	−0.124	−0.082	0.099
y_6：有机电井的村	0.931	−0.120	−0.092	−0.086	0.273
y_7：有灌溉用水塘和水库村	0.045	0.844	−0.024	0.303	−0.310
y_8：有排灌站的村	−0.070	0.080	0.081	0.807	0.180
y_9：来源于国家的村	−0.534	−0.580	0.066	0.186	−0.261

续表

指　　标	F_1 基本生产用能	F_2 水利建设	F_3 劳力用能	F_4 能源基础设施	F_5 能源保障
y_{10}：来源于集体的村	0.436	−0.142	−0.264	−0.013	0.800
y_{11}：来源于其他的村	0.461	−0.689	0.159	0.057	−0.047
y_{12}：无资金的村	−0.372	0.599	0.155	−0.078	−0.611
y_{13}：乡、村办水电站个数	−0.085	−0.366	−0.399	0.710	−0.088
y_{14}：乡、村办水电站发电量	0.053	−0.107	0.041	0.899	−0.041
y_{15}：大中型拖拉机	0.819	−0.365	0.159	0.394	−0.045
y_{16}：小型拖拉机	0.721	−0.059	0.533	−0.079	−0.111
y_{17}：节水灌溉机械	0.273	0.864	−0.002	−0.234	0.073
y_{18}：农用运输车辆	0.557	−0.113	−0.738	0.043	0.125
y_{19}：渔业机动船	−0.101	0.950	−0.168	−0.078	−0.077
y_{20}：联合收割机	0.738	−0.430	−0.089	0.417	−0.068
y_{21}：农村用电量	0.395	0.127	0.014	0.054	0.705
y_{22}：役畜	−0.070	−0.028	0.955	−0.036	−0.028
y_{23}：牛	−0.082	−0.121	0.880	0.200	−0.029
y_{24}：马	−0.273	−0.359	0.658	−0.084	−0.343
y_{25}：驴	0.027	−0.057	0.629	−0.381	0.600
y_{26}：骡	−0.765	−0.362	0.423	−0.056	0.131

第二类：y_7—有灌溉用水塘和水库的村比重；y_9—水利投资的资金主要来源于国家的村比重；y_{11}—水利投资的资金主要来源于其他的村比重；y_{12}—水利投资无资金的村比重；y_{17}—每百户节水灌溉机械套数；y_{19}—每百户渔业机动船艘数。这个因子主要反映全省各个地区的农业生产中的水利建设状况，因此取名为水利建设。

第三类：y_{18}—每百户农用运输车辆数；y_{22}—每百户役畜头数；

y_{23}—每百户牛头数；y_{24}—百户马头数；y_{25}—每百户驴头数。这个因子主要反映全省各地区使用车辆运输、畜力作为劳力状况，因此取名为劳力用能。

第四类：y_8—有排灌站的村比重；y_{13}—乡、村办水电站个数；y_{14}—乡、村办水电站发电量。这个因子主要反映各地区高技术能源基础设施的建设状况，因此取名为能源基础设施。

第五类：y_1—通电的村比重；y_{10}—水利投资资金来源于集体的村；y_{21}—每百户农村用电量（万 kW·h）。这个因子主要包含了电网建设状况及水利建设资金较直接的来源，反映了全省各地区的能源建设现有条件对农民生产用能的保障力度，因此取名为能源保障。

4. 因子得分计算

同样，结合表 3-9 利用与计算各地区农民生活能源使用状况的因子得分函数相类似的方法得出反映全省各地区农民生产用能状况的 5 个公共因子的得分函数为

$$\begin{cases} F_1 = 0.1129y_1 - 0.0333y_2 + 0.1078y_3 + \cdots - 0.1071y_{26} \\ F_2 = -0.0008y_1 + 0.1013y_2 - 0.024y_3 + \cdots - 0.0658y_{26} \\ F_3 = 0.0199y_1 + 0.0321y_2 - 0.0088y_3 + \cdots + 0.0771y_{26} \\ F_4 = -0.0008y_1 + 0.1639y_2 - 0.0732y_3 + \cdots - 0.0213y_{26} \\ F_5 = -0.2738y_1 - 0.0902y_2 + 0.0223y_3 + \cdots + 0.0863y_{26} \end{cases} \tag{3-11}$$

这 5 个公共因子的旋转方差贡献率分别为 29.942%、17.982%、15.2%、12.024%、11.686%，因此反映河北省各地区农民生产能源使用状况的 5 个公共因子的权重依次分别为 0.344 82、0.207 08、0.175 05、0.138 47、0.134 58。同样得到 5 个公共因子和综合因子的得分及排名如表 3-10 所示。

表 3-9　旋转后的因子得分系数矩阵

指　　标	F_1 基本生产用能	F_2 水利建设	F_3 劳力用能	F_4 能源基础设施	F_5 能源保障
y_1：通电的村	0.1129	−0.0008	0.0199	−0.1320	−0.2738
y_2：灌溉主要水源是地表水	−0.0333	0.1013	0.0321	0.1639	−0.0902
y_3：灌溉主要水源是地下水	0.1078	−0.0240	−0.0088	−0.0732	0.0223
y_4：灌溉无水源的村	−0.1345	−0.0471	−0.0131	−0.0245	0.0406
y_5：正常年景用水有保障	0.1260	0.0494	0.0072	−0.0063	−0.0146
y_6：有机电井的村	0.1137	−0.0134	0.0080	−0.0151	0.0321
y_7：有灌溉用水塘和水库村	0.0328	0.1973	0.0354	0.1364	−0.0470
y_8：有排灌站的村	−0.0094	0.0842	0.0696	0.2982	0.1328
y_9：来源于国家的村	−0.0611	−0.1455	−0.0291	0.0104	−0.0952
y_{10}：来源于集体的村	0.0048	0.0114	−0.0320	0.0280	0.2628
y_{11}：来源于其他的村	0.0714	−0.1537	0.0287	−0.0127	−0.0836
y_{12}：无资金的村	−0.0088	0.1007	0.0289	−0.0244	−0.1714
y_{13}：乡、村办水电站个数	−0.0156	−0.0692	−0.1007	0.1978	−0.0274
y_{14}：乡、村办水电站发电量	0.0201	0.0287	0.0492	0.3041	0.0306
y_{15}：大中型拖拉机	0.1287	−0.0536	0.0720	0.1270	−0.0615
y_{16}：小型拖拉机	0.1282	0.0027	0.1613	−0.0050	−0.0734
y_{17}：节水灌溉机械	0.0381	0.1984	0.0408	−0.0209	0.0572
y_{18}：农用运输车辆	0.0477	−0.0464	−0.1836	−0.0135	−0.0211
y_{19}：渔业机动船	−0.0107	0.2082	−0.0127	0.0174	0.0306
y_{20}：联合收割机	0.1096	−0.0813	−0.0026	0.1171	−0.0789
y_{21}：农村用电量	0.0165	0.0856	0.0554	0.0746	0.2629
y_{22}：役畜	0.0236	0.0331	0.2583	0.0285	0.0292
y_{23}：牛	0.0212	0.0216	0.2438	0.1008	0.0344
y_{24}：马	0.0005	−0.0857	0.1381	−0.0438	−0.1192
y_{25}：驴	−0.0152	0.0342	0.1797	−0.0681	0.2304
y_{26}：骡	−0.1071	−0.0658	0.0771	−0.0213	0.0863

表 3-10　公共因子得分情况

地区	F_1 基本生产用能	排名	F_2 水利建设	排名	F_3 劳力用能	排名	F_4 能源基础设施	排名	F_5 能源保障	排名	综合得分	综合排名
石家庄	0.4872	5	−0.1	4	0.2188	5	2.7258	1	0.055	4	0.57	1
唐山	0.0314	8	0.8572	2	−0.015	7	0.1677	3	2.0887	1	0.49	3
秦皇岛	−0.418	9	2.5441	1	−0.586	9	−0.24	7	−0.911	10	0.124	6
邯郸	0.0787	7	−1.122	11	−1.154	10	0.1639	4	−0.211	6	−0.413	8
邢台	0.6627	3	−0.262	5	0.1687	6	0.1934	2	−0.292	7	0.1913	5
保定	0.1618	6	−0.695	10	−1.852	11	−0.409	9	−0.594	9	−0.549	9
张家口	−2.258	11	−0.45	8	0.5258	3	−0.11	5	0.8997	3	−0.674	11
承德	−1.286	10	−0.55	9	0.897	2	−0.145	6	−1.428	11	−0.613	10
沧州	0.7472	2	0.4739	3	0.3793	4	−0.322	8	−0.507	8	0.309	4
廊坊	0.583	4	−0.363	7	−0.427	8	−1.194	11	1.0865	2	0.032	7
衡水	1.2101	1	−0.335	6	1.8461	1	−0.832	10	−0.186	5	0.531	2

第 4 章　河北省地区农民能源使用差异分析

为了更客观、准确、深刻地分析河北省农民能源使用差异状况，并分析其存在差异的原因，我们以公共因子得分情况表为基础，结合河北省第二次农业普查数据及各市的市情等要素进行综合分析。

4.1　生活能源使用差异及原因分析

由表 3-5 可以发现，河北省各地区农民在生活能源的使用上存在很大差异。

比较 F_1 耐用消费品用能因子，位居前三的市是廊坊、沧州、衡水，排名较后的是张家口、承德、秦皇岛。其中廊坊市以 1.482 的得分居全省第一，这说明廊坊市农村耐用消费品耗能最高。耐用消费品耗能与耐用消费品的数量呈正相关关系，即耐用消费品数量越多，耗能也越大。地处北京与天津分界处的廊坊市分别以 108％、12.3％、91.3％、3.3％的农户彩电、汽车、手机、计算机拥有比重，以 2.2％的空调取暖率，95.4％的入户管道水比重居全省第一，同时负相关指标——获取饮用水困难的农户比重为全省最低，仅 1.2％，低于全省平均水平 3.5％。位于河北省西北方向的张家口市分别仅以 87.2％、3.2％、53.4％、0.7％，低于全省平均水平 11.5％、2.3％、21.9％、0.8％的彩电、汽车、手机、计算机拥有比重；分别以 0.083％、6.62％，低于全省平均水平 0.9、28.65％的空调、暖气取暖率列于全省最后。尤其获取饮用水困难的农户比重高达 10.2％，高出全省平均水平达 5.5％，居全省之巅。同时，这与农民的收入水平密不可分，收入越高，农民会越注重生活质量的

改善。廊坊市 2006 年农民人均纯收入达 4950 元，以 205 元的差距仅次于唐山市居全省第二，再受大都市的消费影响，因此廊坊农户绝大部分耐用消费品比重居全省第一。而张家口市农民人均纯收入位于全省最后，仅 2670 元，其对耐用消费品的消费量自然很低。唐山市尽管农村居民人均纯收入居全省第一，但与大都市的距离较远，所受北京、天津的对耐用消费品的需求影响不及廊坊市，其次尽管唐山市农户获取饮用水困难比重为 1.3%，但使用入水管道水户数比重仅 56.3%，低于全省平均值 12.1%，因此排名相对落后，处于全省第五。

比较 F_2 待改善能源因子，排名靠前的三个市分别是邯郸、邢台、石家庄；排名靠后的是唐山、承德、秦皇岛。其中邯郸市以 1.868 的得分居全省第一，这意味着邯郸市农民生活能源消费状况最需要加大力度改善。从普查数据看，邯郸市目前有 14.9 万户，占全市 9.06%的农户没有取暖设施，占全省该类农户数比重达 47.35%；有 4.98 万户，占全市 3.03%的农户饮用水源为雨水和池塘水，占全省该类农户数比重达 54.36%；有 153.39 万户，占全市 93.27%的农户主要炊事用能为煤，占全省该类农户数比重达 25.07%。唐山市则以－1.091 的得分体现了其农民生活能源消费状况为全省最好。由数据显示，唐山市 150.48 万户农户中，有 22.46 万户，占 14.93%的农户主要炊事能源使用煤气、天然气，而全省主要炊事能源使用煤气、天然气的 82.98 万户农户中，唐山市便占了 27.1%。同时，全市仅剩 213 户，占 0.014%的农户饮用水为池塘水和雨水；732 户，占 0.049%的农户无取暖设施。另外，尽管石家庄作为省会城市，但待改善因子却以 1.052 的得分居全省第三，尤其全市仍有 3.3 万户，占 1.9%的农民目前仍无取暖设施；有 1.2 万户，占 13.05%的农民仍饮用池塘水和雨水，这说明石家庄需要加强对农民生活能源使用条件和状况的改善。

由 F_3 健康水资源因子可以看到，廊坊市有 32.71 万户，占 42.13%的农民饮用净化处理水资源，有 44.32 万户，占 57.1%的农民饮用深井水。其中饮用净化处理水资源的农户比重超出全省该指标平均水平 28.5 个百分点，居全省第一。张家口和秦皇岛得分排名紧随其后。衡水市尽管有 91.68%的农户饮用深井水，并且该比重为全省最

高，但因为饮用净化处理水资源的农户比重为全省最低，仅1.37%，最终衡水得分为全省最低，为－1.758。

由F_4沼气因子和F_5太阳能因子可以看出，石家庄市和唐山市新能源的推广是全省较普遍的。其中石家庄使用沼气作为炊事能源的农户为5.88万户，使用太阳能作为炊事能源的农户为0.149万户。唐山市使用沼气作为炊事能源的农户也达5.54万户。

最后看综合得分及综合排名，综合排名前三的是廊坊、唐山、沧州；排名后三位的是邯郸、张家口和邢台。廊坊市尽管F_2待改善能源因子、F_4沼气因子、F_5太阳能因子排名都靠后，但是权重最大(0.337 08)的F_1耐用消费品因子和权重较大(0.127 77)的F_3健康水资源因子均排名第一，综合排名仍居全省第一。张家口市尽管F_3健康水资源因子和F_5太阳能因子得分较高，均排名全省第二，但因为这两个因子权重相对偏小，故综合排名居全省倒数第二。秦皇岛市尽管F_1耐用消费品用能因子得分排名居全省第九，F_3健康水资源因子得分排名居全省第七，F_5太阳能因子得分排名居全省第10，但因为权重相对偏大(0.328 34)的负相关因子——F_2待改善能源因子得分排名全省第九，故综合得分排名仍居第五，排在前列。

4.2 生产能源使用差异及原因分析

由表3-10可以看到，河北省各地区农民在生产能源的使用上也存在很大差异。

比较F_1基本生产用能因子，位居全省前三的是衡水、沧州、邢台，而排名后三的是张家口、承德和秦皇岛。其中衡水市以1.2101的得分居全省第一，从普查可以看出，衡水市4990个村中，使用地下水作为灌溉水源的就有4624个，比重达92.7%，居全省第一；正常年景下用水有保障的村为4772个，比重达95.6%，居全省第二；而灌溉无水源的村个数为15个，仅占0.3%，为全省最少。93.64万户农户中，拥有小型拖拉机的农户达53.195万户，比重达58.81%。同时在全省拥有小型拖拉机的农户中，衡水市的比重达19.525%。灌溉水源充沛，生产

用机械工具普及，这说明衡水市基本生产用能状况为全省最好。而张家口市的 4178 个村中，只有 2322 个，仅占 55.6%的村在正常年景下用水是有保障的；无灌溉水源的村达 1418 个，占全市总村数的33.9%，占全省无灌溉水源村的比重达 41.4%。全市农村拥有大中型及小型拖拉机、联合收割机等农业机械工具数量仅分别为 0.155 万台、5.77 万台、0.0247 万台，无论从总量还是地区每百户拥有量均居全省各市之末。

比较 F_2 水利建设因子，秦皇岛、唐山、沧州分别以 2.5441、0.8572、0.4739 的得分依次位居全省前三位；邯郸、保定、承德分别以－1.122、－0.695、－0.55 的得分依次列于全省后三名。其中秦皇岛市有其优越的先天条件：全市流域面积大于 500km^2 河流 6 条，大于 100km^2 河流 23 条，大于 30km^2 河流 54 条，同时各类水库总库容达 14.86 亿 m^3。因此全市农村的 2269 个村中，有 561 个，占 24.7%的村有能够使用灌溉用水塘和水库，居全省第一。更重要的是，因为其得天独厚的水资源禀赋，农村水利建设对资金的依赖度也为全省最低。普查数据显示，秦皇岛市总共只有 10.9%的村需要各种途径的水利建设资金，有 89.1%的村水利建设不依赖资金。同时，全市每百户农户拥有节水灌溉机械 3.34 台，为全省最高。而邯郸市仅 5.7%的村有能够使用的灌溉用水塘和水库，水利建设有资金依赖的村个数达 32.3%，居全省前列。

比较 F_3 劳力用能因子，衡水、承德、张家口市分别以 1.8461、0.897、0.5258 的得分依次居全省前一、二、三名；保定、邯郸、秦皇岛分别以－1.852、－1.154、－0.586 的得分依次排在全省后三。其中，衡水以平均每百户拥有 45 头役畜、97 头牛、4.3 匹马、11 头驴、11.2 辆农用运输车居全省第一；而保定尽管每百户拥有 30.8 辆农用运输车，但是其役畜、牛、马、驴的百户拥有量仅分别为 0.033 头、0.2 头、0.004 匹、0.035 头，处于全省最后一名。

比较 F_4 能源基础设施因子，石家庄作为省会城市以 2.7258 的得分居全省第一，邢台、唐山分别以 0.1934、0.1677 的得分位列全省第二、第三；廊坊、衡水、保定分别以－1.194、－0.832、－0.409 的得分居

于全省后三位。其中石家庄的4450个村中，有594个，占13.4%的村具有排灌站，并拥有31个发电量为3106.91万kW·h的乡、村办水电站，居全省第一。邢台尽管没有乡、村办水电站，但5163个村中，有636个，占12.3%的村具有排灌站，因此基础设施因子得分排名为全省第二。唐山尽管拥有7个发电量为10万kW·h的发电站，但是因为其拥有排灌站的村比重仅9.4%而位居第三。廊坊市不但没有乡、村办水电站，拥有排灌站的村比重也仅为3.1%，因此排名最后。

比较F_5能源保障因子，可以发现，唐山有27.4%的村可以从集体筹集水利建设资金，同时农村户均用电量能达0.686万kW·h，这样唐山以2.0887的得分居全省第一，说明唐山市农民在生产用能方面最有保障。承德市以－1.428的得分居全省最后。前面分析了秦皇岛因水资源充沛，水利建设对资金的依赖度较低，故水利建设因子排名第一。但因为其农民生产能源的资金来源保障不足，致使秦皇岛市在能源保障因子上排名倒数第二。

再来看综合得分及排名，石家庄尽管只有F_4能源基础设施因子排第一，其余因子均排名第四或第五，但因为F_4得分非常高，达2.7258，而其余因子得分最高仅0.4872，最终综合得分为0.57，排名全省第一。这说明石家庄作为河北省省会，农民生产能源使用状况发展均衡。综合排名前三的还有衡水(0.531)和唐山(0.49)。排在后三的分别是张家口(－0.674)、承德(－0.613)、保定(－0.549)。

综上，我们可以清楚地看到，河北省11个地区无论是在生活用能方面还是在生产用能方面均存在很大差异。不仅有能源消费结构上的差异，也有能源消费能力上的差异，还有能源消费条件上的差异。这些差异既有来自地区间的经济发展状况的因素，也有来自地理位置的因素。这些差异的存在，在很大程度上阻碍了河北省新农村建设的平稳、健康、有序开展，也加大了河北省解决“三农”问题的难度，更对河北省经济、社会和谐发展构成了一定负面影响。

第 2 篇

加快新农村建设
提高农民满意度

第5章　新农村(民居)建设的概念与理论概述

5.1　新农村(民居)建设的相关概念

1. 社会主义新农村

“社会主义新农村”这一概念早在20世纪50年代就提出过。20世纪80年代初，我国提出“小康社会”概念，其中建设社会主义新农村就是小康社会的重要内容之一。此次五中全会所提建设“社会主义新农村”，则是在新的历史背景中，在全新理念指导下的一次农村综合变革的新起点。对于这次中央提出建设社会主义新农村与以往的新农村建设相比有何新特点，中国人民大学农业与农村发展学院院长温铁军指出，新农村新在：其一，改变以往简单化的加快城市化的倾向，更关注城乡之间良性互动；其二，农村的发展能够体现科学发展观的要求，重在农村社会相关制度建设；其三，农村发展仍然保持田园风光，是人与自然环境令人耳目一新的良好生态社会。十六届五中全会提出的建设社会主义新农村，新在农村建设有一个全新的环境和背景，亦即我国已进入“以工补农，以城带乡”的新阶段；新在发展农村的文化教育事业，培养有文化、懂技术、会经营的新型农民，提高农民的整体素质，给农民创造一个平等的自我发展空间。

2. 人居环境

“人居环境”一词最早出现于希腊著名的城市规划专家道萨迪亚斯(Doxiadis)1968年所著的《人类聚居学》一书。“settlement”一词一般译为“聚居地”或“村落”，包括城镇。我国中科院院士吴良墉将其译为“聚居”。20世纪80年代，他在研究道氏“人类聚居学”的基础上，结合

中国的具体国情,提出了“人居环境科学”。

道氏的“人类聚居”是一个内容十分宽泛的概念,他认为:“人类聚居是人类为了自身的生活而使用或建造的任何类型的场所”和“人类聚居是人类生活其间的聚居”,“人类聚居实际上指的是我们的生活系统”。道氏认为,人类聚居是地球上可以供人类直接使用的、任何形式的、有形的实体环境,人类聚居不仅是有形的聚落本身,也包括了聚落周围的自然环境;人类聚居还包括人类及其活动,以及由人类及其活动所构成的社会。

吴良墉认为人居环境是人类聚居生活的地方,是与人类生存活动密切相关的地表空间,它是人类在自然界中赖以生存的基地,是人类利用自然、改造自然的主要场所。他认为:“人居环境的核心是‘人’,人居环境研究以满足‘人类居住’需要为目的。”“大自然是人居环境的基础,人的生活、生产以及具体的人居环境建设活动都离不开更为广阔的自然背景。”“人居环境是人类与自然之间发生联系和作用的中介,人居环境建设本身就是人与自然相互联系和作用的一种形式,理想的人居环境是人与自然的和谐统一,或如古语所云‘天人合一’。”“人创造人居环境,人居环境又对人的行为产生影响。”

3. 民居

民居通常被定义为“本土的”、“自发的”,由本地居民参与的适应自然环境和基本功能的营造。《中国大百科全书》将民居定义为“宫殿、官署以外的居住建筑”。今天,中国建筑界的大多数专家和学者都不同意将民居的概念仅局限于住宅,而是认为它的内涵应该扩大到城镇和村落中与生活相关的各类建筑,甚至是聚落本身。大体而言,“民居”被界定为非官方的、非专家现象的限于日常生活领域的人类居住环境。本章中为研究而使用的“民居”一词,亦更是限定民居的范围,集中于土生土长的乡村之中,伴随着农耕文化的发展而形成的村寨、乡镇聚落和住宅。

4.“二十字方针”

“二十字方针”指中央对我国新农村建设提出的“生产发展、生活富

裕、乡风文明、村容整洁、管理民主”的建设要求。

5.2　新农村(民居)建设的相关理论

1. 人地关系理论

所谓人地关系理论是指各种人文现象与自然环境的关系在人们头脑中的反映，即人们对自然环境在文化(包括经济)发展中作用的看法。它是人类文化系统的重要因素，是文化价值观的核心内容之一，对人类利用和改造自然环境的方式和强度有着重要的影响和作用。我国中科院院士吴传钧先生将人地关系思想完整地引入地理学中，提出和论证了人地关系的地域系统是地理学的研究核心。他在研究法国人地学派代表人物 delaBlache V 和 Brunhes J 的“或然论”的基础上，经过长期的研究和探索，提出“人”和“地”这两方面的要素按照一定的规律相互交织在一起，构成人地关系地域系统。他还提出，人地关系地域立体研究的中心目标是协调人地关系，从空间结构、时间过程、组织序变、整体效应、协同互补等方面去认识和寻求全球的、全国的或区域的人地关系系统的整体优化、综合平衡及有效控制的机理，为有效地进行区域开发和管理提供理论依据。在人地关系的实际研究中，可通过最能体现人地关系本质的关键要素(联结点)来剖析其主要问题。从人地关系发展阶段看，最初是通过粮食、居所、资源和交通等基本的“联结点”来实现人地联系，所以以往人地关系的研究多侧重于人类生存所依赖的自然资源的开发利用和生产建设，而对人类生活，如居住环境状况研究得不够。居所是人类休息、娱乐等活动的主要场所，是保证人类劳动力再生产的重要基地。住房与住区也是农村生产、经济、社会等活动的后勤保障基地。因此，居住环境的好坏直接影响到人类生活，从这个意义上讲，人居现象应该说是人地关系最基本的联结点，农村民居建设的研究应以人地关系理论为指导，努力实现人地协调。

2. 可持续发展理论

1987 年，世界环境和发展委员会(WECD)在《我们共同的未来》中指出，可持续发展是既满足当代人的需求，又不损害后代人满足其需求

的能力。这一概念包含三层含义。一是“需要”的含义，既要满足当代人的需要，又不危害后代人的利益；二是“限制”含义，人类对资源与环境的索求不应超出它们的承载极限；三是“协调”的含义，当代与后代、局部与整体、自然与社会、发达地区与落后地区等，强调“适度”和“协调”前提下的公平发展。从系统论的角度理解，可持续发展是指以人为中心的“自然-经济-社会”复合系统的可持续，即生态持续发展、经济持续发展和社会持续发展。其中生态持续发展是基础，经济持续发展是条件，社会持续发展是目的。由此可见，可持续发展作为一种新的发展观，是以人类的永久生存和发展为核心，以资源和环境的承载力为限度，以人的全面发展和社会的全面进步为目标的持续过程，即是以人为本的。它是具有高度战略意义和实践意义的指导思想，已引起各国政府高度重视和认可，其“以人为本”的目标理论正是新农村建设的理论支撑。

3. 城乡一体化理论

工业革命以后，西方城市高速发展，城市发达与乡村贫困的反差越来越大，城乡之间的矛盾与冲突越演越烈。为解决这一严重的城乡矛盾，早期的空想社会主义学者提出城乡一体化理论的原始构想，如莫尔提出了“乌托邦”，傅里叶提出了“法郎吉”模型，欧文提出了“新村”等。在这个基础上，19 世纪末，英国社会活动家霍华德提出了“田园城市”社会改良方案，1892 年，恩温 (Runwin)在规划建设基础上进一步发展了该理论，并把它付诸城市区域发展实践。20 世纪 60 年代，美国著名的城市文化学家刘易斯·芒福德从保护人居系统中的自然环境出发，强调乡村在城乡关联发展中的特殊重要性，他指出：“城与乡，不能截然分开；城与乡，同等重要；城与乡，应当有机结合在一起。”20 世纪 80 年代以前的城乡发展理论具有共同的观点取向，就是认定城市与乡村具有严格界限与差别，而且城乡差别将永远存在。但是到了 20 世纪 80 年代，加拿大学者麦基(Maggie T G)通过亚洲许多国家的实证得出，第三世界国家走的是不同于西方传统的城市化道路：在实行工业化的同时，没有发生人口大量迁入城市的现象，而是形成了“城市与乡村界限日趋模糊，农业活动与非农业活动紧密联系，城市用地与农村用

地相互混杂的结构”。麦基基于这一理论设计出了著名的 Desakotasi 模型,表达出一种非城非乡,但又具有城乡双重特征的城乡一体化区域模式。我国学者在借鉴国外理论的基础上,根据我国的实际情况提出了我国的城乡一体化理论:城市与乡村在政治、经济、文化等方面的广泛融合和城乡发展的有机结合,形成“以城带乡,以乡促城,优势互补,共同发展”的城乡关系,达到城市现代化、乡村城镇化,使城乡融为一体,使区域经济结构及社会分工合理化,综合协调发展。

4. 农村城镇化理论

农村城镇化是指各种要素不断在农村城镇中集聚,农村城镇人口不断增多,城镇数量、规模不断增大,质量不断提高的过程。它是以工业为主体的非农产业集聚发展的必然结果,是农村社会演进并通往现代化的一个重要过程,是传统农村向现代城市文明的一种变迁,是统筹城乡发展、全面建成小康社会的重要内容。城镇化的本质特征主要体现在三个方面:一是农村人口在空间上的转换;二是非农产业向城镇聚集;三是农业劳动力向非农业劳动力转移。我国还存在耕地资源短缺,且耕地质量较差的问题,因此人均占有粮食水平并不宽余。从养活中国人来说,目前的耕地数量是基本可以的,但需要花大的气力来维持粮食种植面积的稳定和产量的提高,以求解决“吃饭”和“建设”问题的平衡。农村城镇化是实现我国农村经济发展的大战略,是我国实现小康社会的必由之路。我国人口尚在不断增加,同时由于经济社会的综合发展,我国城镇化水平还在不断提高,城镇数量还要增加,城镇规模也要拓展。据有关方面的预测,我国人口高峰年大约在 2030 年,人口总量将达到 16 亿。因此,我国尚需要增加城镇空间以满足人口增加的需要。这就需要进行农村建设用地的整理,通过农村民居的改造和建设,使新农村建设逐步集中、集约,不仅可以补充耕地资源的不足,还可以为城镇社会经济发展提供宝贵的土地资源,是现阶段我国保持耕地资源总量平衡、促进城镇化发展的重要途径。

5. 新农村建设的相关政策及实施情况

2005 年,我国开始探索和实施新农村建设。2006 年党的十六届

五中全会提出了建设社会主义新农村的重大历史任务，明确了建设“生产发展、生活宽裕、乡风文明、村容整洁、管理民主”的新型农村社会的目标。十六届五中全会党中央明确提出“建设社会主义新农村是我国现代化进程中的重大历史任务”；2006 年 2 月，中共中央、国务院下发了《中共中央国务院关于推进社会主义新农村建设的若干意见》的“一号文件”。由此，新农村建设工作在全国各地蓬勃展开，新农村建设问题成为全社会关注的焦点问题。为了响应中央号召，中共河北省委制定了《河北省人民政府关于推进社会主义新农村建设的指导意见》，指出建设社会主义新农村的指导意见和方向。新民居建设作为社会主义新农村建设的一个重要指标，河北省政府作为一项重点工作进行安排部署。2009 年河北省批准并支持了 1000 个省级新民居示范村的建设，2010 年再启动 6500 个新民居示范村的建设，并着重抓好 2000 个重点示范村的建设。在实施过程中，政府采取许多优惠和补贴政策，如 2009 年河北省安排新民居建设专项资金 2 亿元，重点用于支持示范村新民居建设、村庄规划和配套公共服务设施建设，各市、县(市、区)也安排了专项资金用于推进新民居建设。除此之外，各级城乡规划建设部门组织有关单位和专家成立规划设计、施工建设、新材料推广等若干技术服务小组，进行技术服务和巡回指导，及时解决规划建设中存在的问题。尽管如此，新民居建设并没有得到农民百分之百的称赞，究其原因是多方面的，而且，农民对于相关政策的了解程度和相关措施的满意程度，对于河北省推行新农村建设和新民居建设有很大的影响。

第 6 章 数据挖掘的基本概念

6.1 数据挖掘概述

1. 数据挖掘的概念

数据挖掘，也称为数据库中的知识发现（knowledge discovery in database，KDD），是从数据库中抽取隐含的、以前未知的、具有潜在价值的信息的过程。具体来讲是指从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的，但又是潜在有用的信息和知识的过程，就是利用各种包括数据仓库在内的分析工具在海量数据中发现模型和数据间的关系并做出预测的过程。

2. 数据挖掘的功能

利用数据挖掘技术可以帮助获得决策所需的多种知识。一般地，数据挖掘任务可以分成描述和预测两种类型。描述型数据挖掘的任务是寻找数据集中数据的一般特性，并对数据中存在的规则进行描述，或者根据数据的相似性把数据分组。

预测型数据挖掘的任务是在当前数据上进行推断和预测，并根据数据项的值来确定某种模式。数据挖掘的目标是从数据库中发现隐含的、有意义的知识，主要有以下几类功能。

(1) 分类：其旨在生成一个分类函数或分类模型，该模型能把数据库中的数据项映射到给定类别中的某一个。

(2) 关联分析：关联规则(association rules)有时又被称为购物篮分析方法，以发现交易数据库中不同项之间的联系。

(3) 聚类分析：聚类(clustering)是在事先不规定分组规则的情况

下，将数据按照其自身特征划分成不同的群组。

（4）概念描述：就是对某类对象的内涵进行描述，并概括这类对象的有关特征。

（5）偏差检测：寻找观测结果与参照值之间有意义的差别。

3. 数据挖掘流程

数据挖掘作为 KDD 的一部分，尽管狭义观点认为数据挖掘只是 KDD 的模式发现阶段，但从广义观点看，其可认为就是数据库中的知识发现，包括数据清理与集成、数据选择与变换、模式发现以及模式表示与评估。而且在企业界、媒体及数据库研究领域，已将数据挖掘作为 KDD 的代名词。因此，可以将数据挖掘流程图表示为图 6-1 所示。

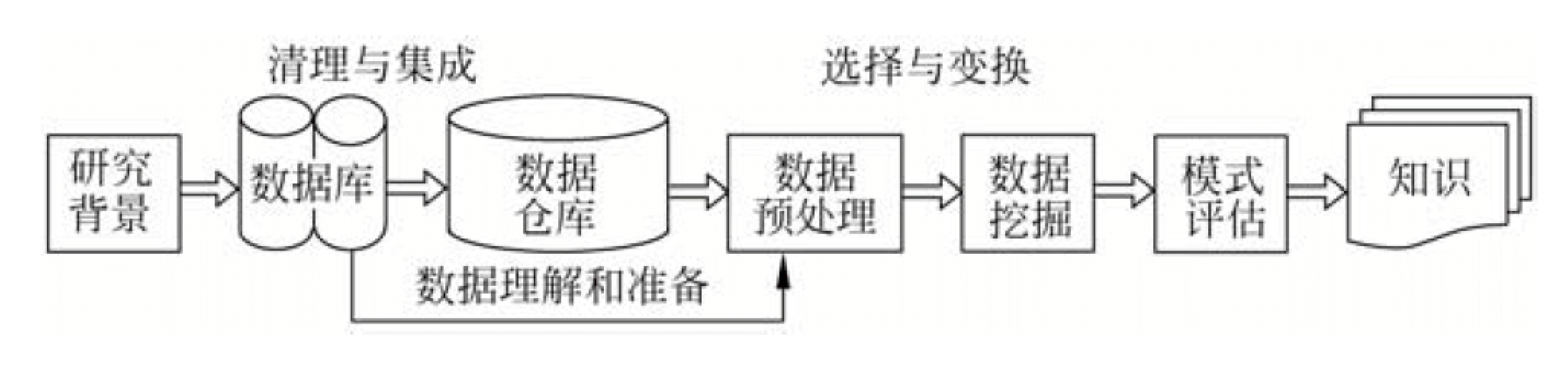


图 6-1　数据挖掘流程图

6.2　数据挖掘在调查数据中的可用性分析

对于数据挖掘技术运用于调查问卷，可结合数据挖掘的功能和问卷数据的特征进行分析。问卷调查，无论是企业所做市场预测，还是社会科学研究所做社会调查，抑或政府所做民意测验，都是在最大可能下掌握被调查者的人口统计特征，如年龄、性别、地域、教育文化水平、可支配收入、职业、消费观念、个人偏好等。这样我们在分析问题时，就可以将具有异质特征的目标人群进行聚类细分，细分后的每一类人群都会具有相似的消费、需求或其他行为特征。这有利于决策部门有针对性地实施产品、技术或政策推广策略。在问卷调查的数据采集、录入过程中难免发生各种错误，这样会导致异常值的产生。在数据挖掘流程中，对异常值的检测和处理能有效地降低调查误差。

各类调查中，都或多或少地涉及多项选择，对于多项选择题目，传

统的处理方法主要是通过频数分析或对应分析等方法来说明问题。而对应分析就是将指标变量的因子分析(R 型因子分析)和样品的因子分析(Q 型因子分析)结合起来进行统计分析,从 R 型因子分析出发,直接获得 Q 型因子分析的结果,并将指标和样品分析的结果同时反映到相同坐标轴的一张二维图形上,对问题进行较为直观的分析。但对应分析也存在一些不足,首先,对应分析不能用于相关关系的假设检验。它虽然可能揭示变量间的联系,但不能说明两个变量之间存在的联系是否显著。其次,在对应分析中所做图形为了直观通常是二维的,这种做法是一种降维的方法,将原始的高维数据按一定规则投影到二维图形上,投影后的分类大多数情况与原始数据大体一致,但也可能与原始高维数据的分类有较大的差异情况存在。再有,对极端值敏感。在分析过程中,极端值对对应分析的结果影响较大。在进行分析之前,要将列联表中的数据审视,避免极端值的存在。比如,列联表数据应为正值,若有为零的数据,可视情况将相邻的两个状态合并。我们可以将多选题类比为购物篮,采用关联规则,不仅可对多选题候选项进行分析,也能将多选题候选项与其他变量进行交叉分析。关联规则挖掘不仅可以避免受极端值的影响,还能通过关联规则中的支持度、置信度或作用度等各指标值来确定变量之间的关联强度和相关程度。

第7章 新农村建设的调查过程及数据准备

7.1 新农村建设的调查方案设计

1. 确定调查目的

此次调查的目的在于全面了解河北省农村居民对于新农村建设和新民居建设各个方面的满意程度及对当前农村发展中所关注的各项问题。建设社会主义新农村是构建和谐社会的基础，是统筹城乡发展，全面实现小康社会的重要途径。为了解当前新农村建设的现状，为河北省委、省政府指导“三农”工作尤其是新民居建设提供决策依据，所以组织进行了此次新农村建设情况调查。

2. 参与调研农村居民的确定

本次调研最终挑选河北省石家庄市、邢台市、唐山市、保定市等地区新民居建设试点村(2009年名单)进行问卷发放，从河北省受表彰的推进社会主义新农村建设先进县(市、区)名单中选取石家庄、邢台、唐山、保定四个地区，共选20个示范村，再选20个非示范村作对比调查并对试点村相邻的普通村做出样本量大致相等的对比调查。

3. 样本量和抽样方法的确定

本次调查采用问卷调查的方式进行。样本量的确定综合考虑了地区、人口、经济等多种因素，进行合理的配额设计，以确保样本的抽取尽可能体现河北省农村居民实际情况。调研采用纸质问卷，预计发放1000～1200份。本次调查采用分层随机抽样方法，各个市样本量根据农村人口所占参与调查地区总体农村人口比例确定，但是由于所调查

村庄各项指标数据难以寻求，故经过大致估算，采用每个村庄随机抽取不少于 20 位不多于 30 位村民进行调查，以确保样本村庄的代表性。

4. 抽样调查方式及地点

考虑到农村居民文化水平，为了提高调查问卷回收率，尽可能减少缺失项，所有问卷采用面访的方式进行调查，由调查人记录参与调查的农村居民回答内容填入问卷。对每个村庄中的居民采用简单随机抽样方法，即大致了解村庄结构后，根据样本量估算调查间距，选取农户进行调查。

7.2　新农村建设的问卷设计与指标选取

1. 问卷设计

问卷内容包括农村民居建设模式、农村经济和农民收入、农民生产、生活、居住环境、政策落实等方面，在不影响农民回答问卷积极性的前提下尽可能多地涵盖新农村建设和新民居建设中的各个方面。调研问卷主要由以下几个部分组成：①填卷说明，包括国家新农村政策基本介绍、新民居建设要求等内容；②人口统计学信息，包括所在村庄名称和类型、性别、年龄、文化程度、政治面貌、居住类型；③对农村目前现状的评价，包括对当前居住村庄的收入、社会风气、村干部、居住环境等方面的满意程度；④对新农村和新民居建设的政策理解和建议，主要包括对国家政策的理解、对当前新农村建设的期盼等。具体问卷设置详见附录 A。

2. 指标选取

本研究中选取问卷部分问题进行分析，其中包括：人口统计学变量，用于农村居民人群广度细分；有关农村居民总体和各个方面满意度的变量，其选择依据参照新农村建设“二十字方针”；有关农村居民政策理解和政策支持的变量和有关农村居民自身困难和政策期望的变量，用于关联规则挖掘。参与本次分析的变量共有 91 个，具体变量设置和在问卷中的位置详见以下的分析。

7.3 新农村建设的调查数据录入

对收集上来的问卷利用 Epidata 3.1 进行录入，具体录入方式为：单项选择题按照问项从左至右依次编号为 1、2、3、4、…进行录入；多项选择题若某个选项被选，录入为 1，未选录入为 0。最终整合数据，转化为 Excel 文件。

7.4 新农村建设的调查数据预处理

本研究采用数据挖掘中数据预处理方法，主要处理内容如下：①数据审核，对于缺失值过多的变量（问项）和无效问卷（主要为所有选项均填一项、缺失值过多或者前后逻辑不符等情况）进行删除处理；②使用最可能的值填补空缺值，最可能的值可以根据判定树归纳来确定；③数据分割，对应数据挖掘所关注问题将数据进行分割，以避免流程中节点过多引起的混乱；④变量转换，将用于关联分析的若干变量选项进行合并，以避免某个样本标志个数不够频繁。

本研究主要应用 Clementine 12.0 和 SAS EM 模块对数据进行预处理。数据审核部分将回答问项少于问卷总体问项 50％的问卷视为无效问卷，作删除处理，此类问卷共 136 份，另外剔除无效观测 59 份（回答选项均选一项、回答逻辑前后明显不符等情况），最终得到有效问卷 905 份，其中示范村问卷为 576 份，有效问卷回收率 82.3％。问卷回收率较高，可以认为能够反映出河北省新农村建设和新民居建设情况；删除了部分变量，将其剔除出模型，如涉及“大学生村官问题”的变量，此类变量的共同特点是缺失值过多，或者答案过于单一。由于此次问卷调查为面访式，问卷均由访问者填写，所以几乎不存在问卷数据中涉及人口统计学变量值的缺失问题，对于问卷中有其他少量缺失的变量，文章采用决策树算法对其进行插补，此种缺失值处理方法的优点在于缺失值插补是根据现有的大量数据进行推测得到的，用于推测空缺值最为合理，此结论以及缺失值插补正确率的检验过程在图 7-1 的流

程图中得到验证，具体过程不再赘述①。应用数据分割节点，将处理之后的数据分割成三个部分，分别将用于决策树（CART 算法）分类、关联规则挖掘（GRI 算法）、关联规则挖掘（Apriori 算法）。最后，通过对某些变量的转换，得到适于进行关联分析的变量。

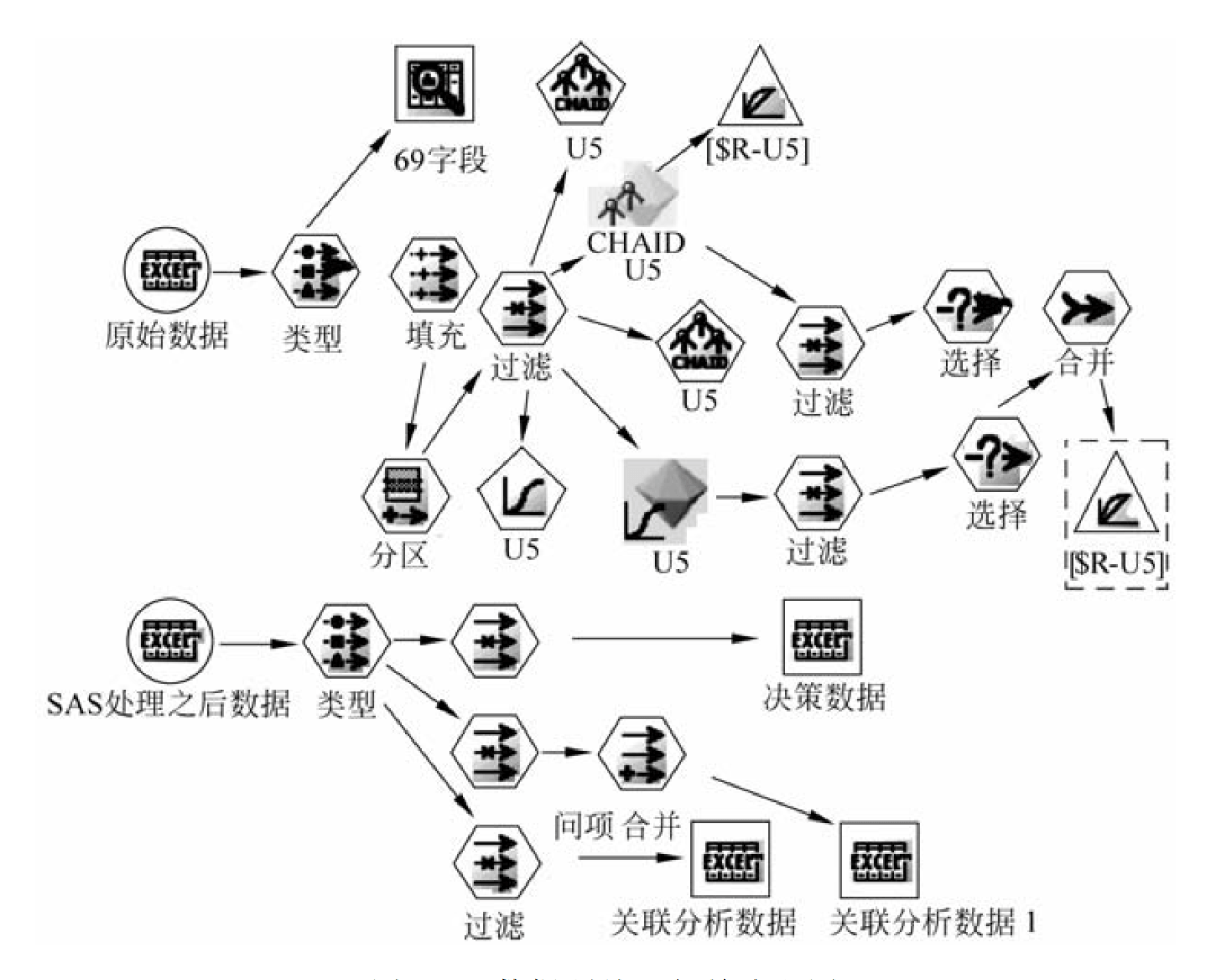


图 7-1　数据预处理相关流程图

① 在本研究的分析中，利用 Clementine 12.0 相关节点对缺失值处理预测正确率进行验证，但是由于算法处理时间过长，本研究中全体数据缺失值处理应用 SAS 9.1 企业数据挖掘模块进行处理并导出，插补方法仍然为决策树（CART）方法。

第 8 章　新农村建设的描述性统计分析

8.1　调查对象基本情况的描述性统计分析

在前面研究的合理性、准确性的前提下，对此次调查结果中示范村的调查问卷按照不同性别、年龄、文化程度、职业、政治面貌和村类型进行汇总，见表 8-1、表 8-2、表 8-3，参与问卷调查的居民中男性和女性的人数分别为 296、280；年龄在 60 岁以上的人数最少，30 岁以下和 30～45 岁的人数分别是 185 和 175；初中学历的人数高达 232，大专及以上学历的人数极少；职业是农民群众的人数为 394，村干部或乡镇干部的人数较少；党员有 73 人，民主党派有 21 人，政治面貌是其他的人数最多，高达 482；新农村示范村有 281 个，移民搬迁村有 110 个，新农村达标村个数较少。

表 8-1　被调查者的性别、年龄分布表

项目类别	性别		年龄			
	男	女	30 岁以下	30～45 岁	45～60 岁	60 岁以上
人数	296	280	185	175	153	63

表 8-2　被调查者的文化程度、职业分布表

项目类别	不识字或少识字	文化程度				职业			
		小学	初中	高中	大专及以上	农民群众	村干部	乡（镇）干部	其他
人数	40	95	232	185	24	394	18	20	144

表 8-3　被调查者的政治面貌、所属村类型分布表

项目类别	政治面貌			村类型			
	中共党员	民主党派	其他	新农村示范村	新农村达标村	移民搬迁村	其他村
人数	73	21	482	281	35	110	150

8.2　居民对所在村(镇)目前的农村现状的评价分析

关于居民对所在村(镇)目前的农村现状的评价,本次调查涉及了很多方面,如农民收入情况、"六通、六有、六化"建设情况、农民生存环境情况、农民教育文化水平情况、农民对各项社会保障制度满意情况、村级班子建设情况等,本节从不同角度选取几个与农民生产生活息息相关的方面进行简单描述性统计分析,具体的选取情况见表 8-4。

表 8-4　居民对农村现状评价的指标选择

方　面	角　度
农民收入	居民对自家家庭收入提高状况的满意度 居民收入的主要来源
"六通、六有、六化"的建设	居民对所在地"六通、六有、六化"建设的满意度
农民生存环境	居民对自身目前的生活状态的满意度 居民对所在村(镇)生存环境的满意度 居民对所在村(镇)娱乐设施的满意度 居民对现在的居住环境的满意度
农民教育文化水平	居民对所在村(镇)的教学条件的满意度 居民对目前农村的教育环境的满意度
社会保障制度、社会治安	居民对所在村(镇)各项社会保障制度的满意度 居民对所在村(镇)的农村社会风气的满意度 居民对所在村(镇)社会治安的满意度
村级班子的建设	居民对所在村(镇)班子干部的满意度 居民最希望大学生村官提供哪些方面的帮助

8.2.1 农民收入的现状分析

新农村建设中，农民收入最能代表农村居民的生活水平，也最能体现新农村建设的程度，而农民收入的来源能体现出农民获得收入的渠道的广泛性，渠道越多说明人们生活就越好，就越接近新农村建设的目标，因此本小节选取收入提高状况和收入来源两个方面来进行描述分析。

从居民对自家家庭收入提高状况的满意度评价的调查结果来看（见图 8-1），56.3%的人对自家家庭收入的提高状况比较满意，9.5%的人对自家家庭收入的提高非常满意，说明这部分人的家庭收入在新农村建设过程中提高了，对自家家庭收入的提高“不太满意”的人数占27.3%，对自家家庭收入的提高状况“不满意”的人数占 6.9%。

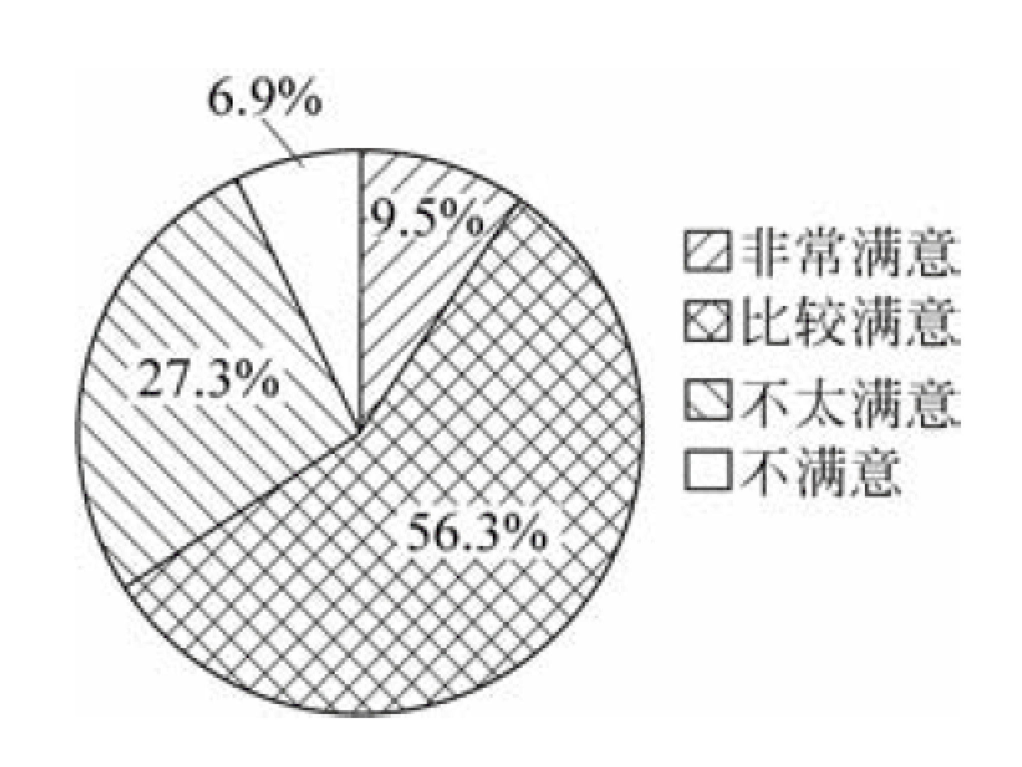


图 8-1　居民对自家家庭收入提高状况的满意度评价

在过去，农民主要是依靠种庄稼、干体力活获得收入，干活累且挣得少，新时代实施新农村建设以来发生了很大变化，从对居民收入的主要来源的调查结果来看（见图 8-2），现在农民主要通过做生意获得收入的比例占 26.4%；而通过农业获得收入的比例占 25.2%，说明现在从商的农民数量大大增加；靠养殖业获得收入的人数较少，仅占2.3%；在乡镇企业上班获得工资收入的人数占 15.1%；外出务工的人占19.0%。

8.2.2 “六通、六有、六化”建设的满意度评价分析

新农村建设的主要项目中，“六通、六有、六化”建设是主要项目，此

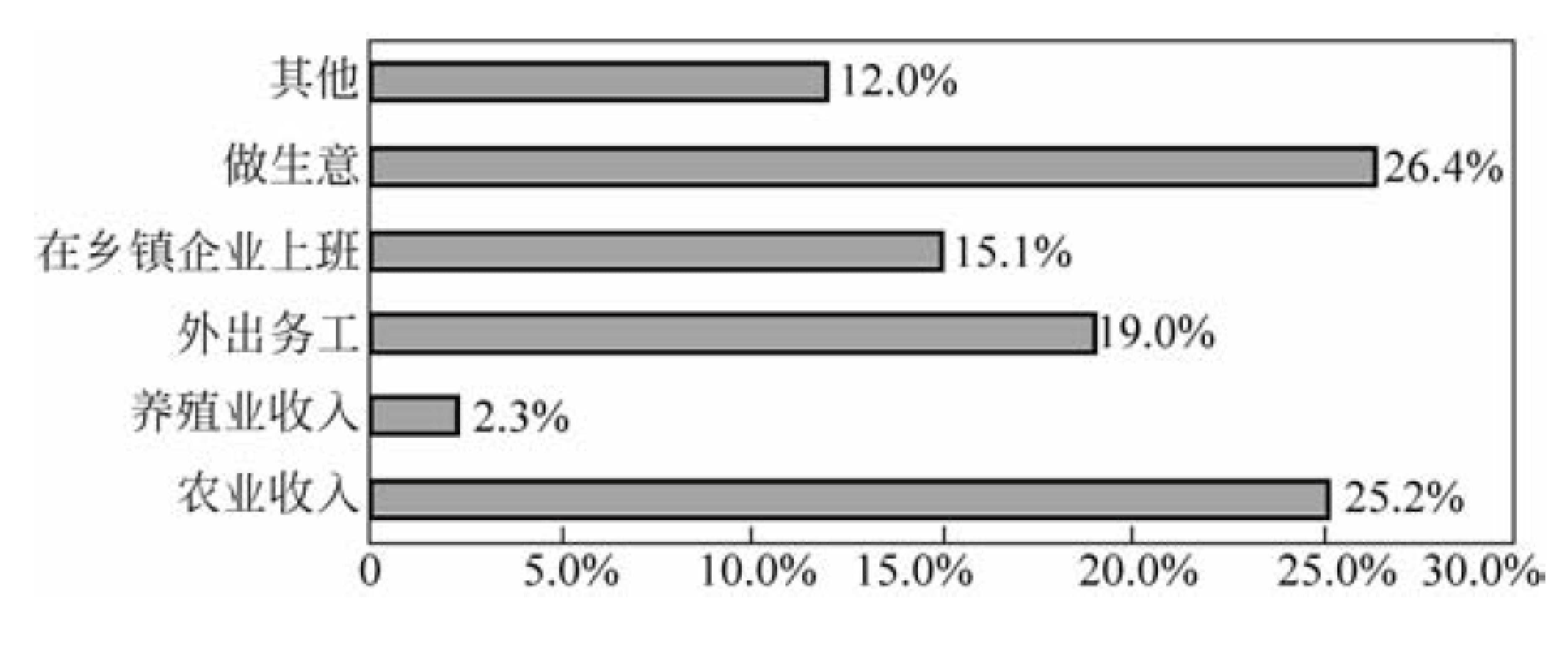


图 8-2 居民收入的主要来源

项目完成的力度能够体现出农民在物质生活和精神生活方面的提高，从对居民所在地“六通、六有、六化”建设的满意度评价结果来看(见图 8-3)，8.2%的人对“六通、六有、六化”建设非常满意，60.9%的人对“六通、六有、六化”建设比较满意，说明新农村“六通、六有、六化”建设取得了显著作用；除此之外还有 23.3%的人对“六通、六有、六化”建设不太满意，7.6%的人对“六通、六有、六化”建设不满意，说明现在的“六通、六有、六化”建设还有再提高的空间，相信随着时代的进步，新农村建设会达到更大的进步。

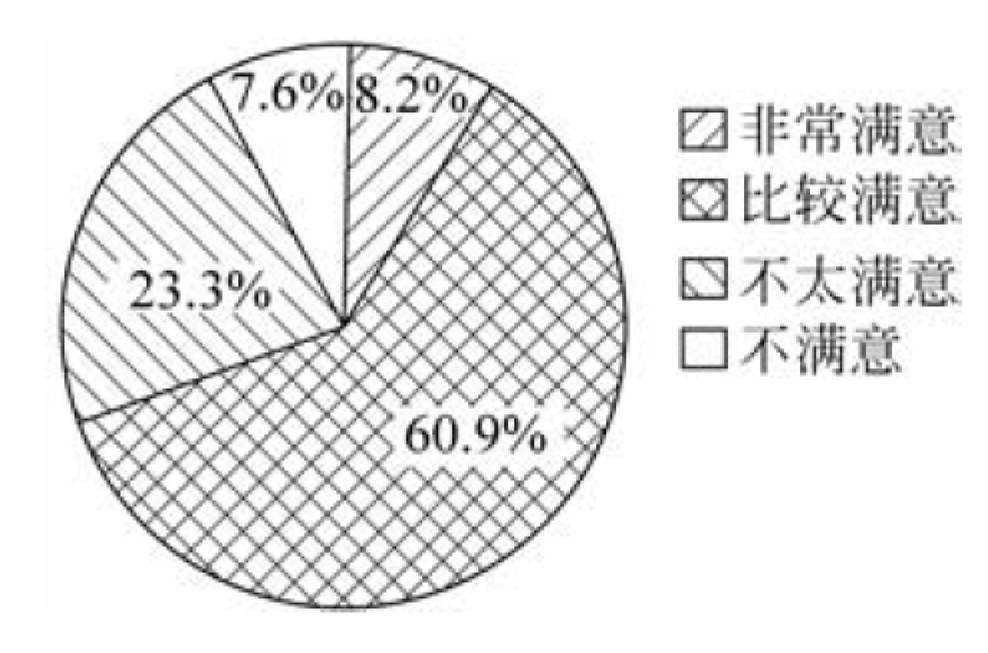


图 8-3 居民对所在地“六通、六有、六化”建设的满意度评价

8.2.3 农民生活生存环境的满意度评价分析

本小节主要针对农民现在的生活和生存的现状进行分析，首先对居民自身目前生活状态进行满意度调查，然后分别对居民所在地的生存环境和居住环境进行满意度调查，其中包括居民对娱乐设施的满意度调查。

居民对自身目前生活状态的满意与否最能表达出在新农村建设的过程中人民生活水平的提高程度，满意度评价高就说明人们享受到了新农村建设带来的好处。从此次调查结果来看(见图 8-4)，分别有

10.1%和65.3%的人对自身目前生活状态非常满意和比较满意，满意度评价较高，然而还有18.2%的人对自身目前生活状态不太满意，6.4%的人对自身目前生活状态不满意，这部分人占少数一部分，新农村建设使得居民对自身生活有了更客观的评价。

从对居民所在地生存环境的满意度评价来看(见图8-5)，总体来说，居民对所在地生存环境的满意度评价较高，其中有10.2%的人对所在地生存环境非常满意，有57.7%的人对所在地生存环境比较满意，说明在新农村建设过程中这部分人的生存环境有很大的提高；而有23.4%的人对所在地生存环境不太满意，8.7%的人对所在地生存环境不满意。

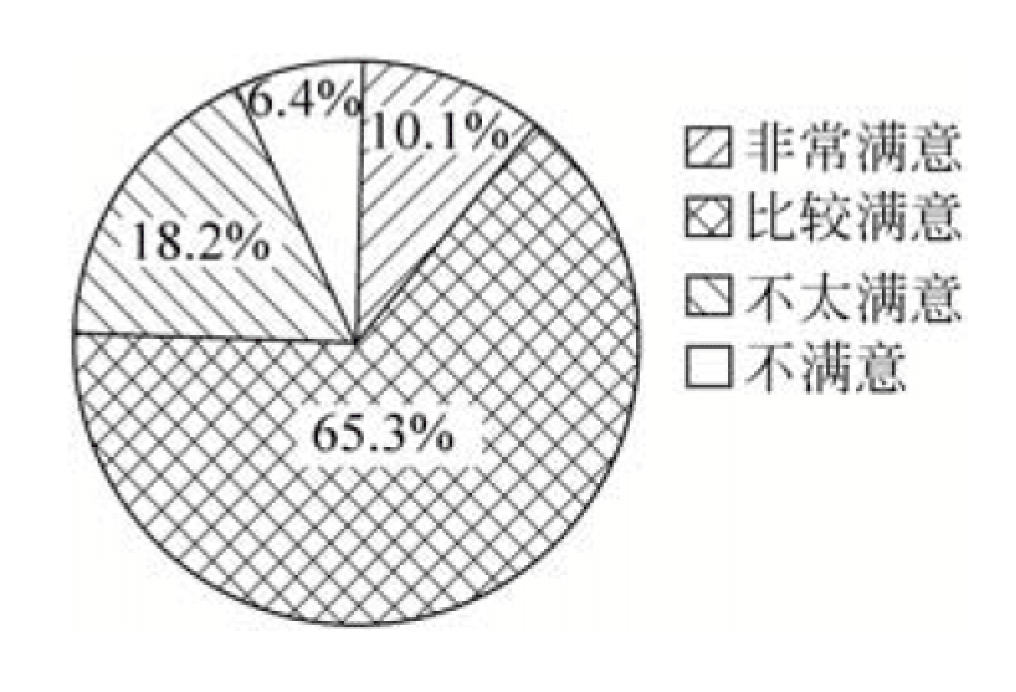


图8-4　居民对自身目前的生活状态的满意度评价

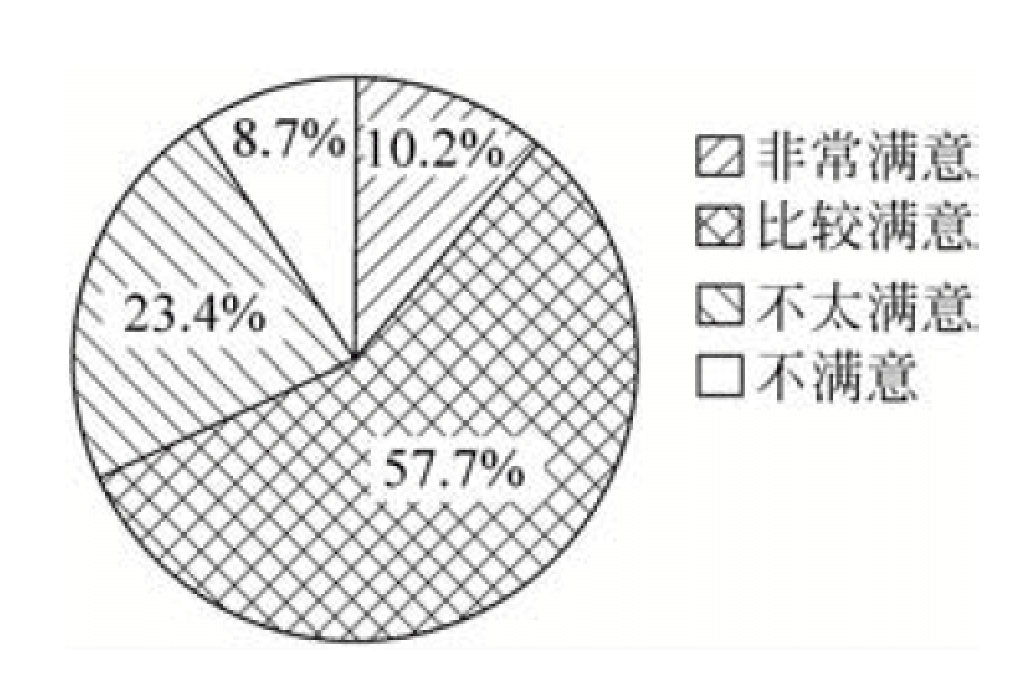


图8-5　居民对所在地生存环境的满意度评价

新农村建设的主要内容中对娱乐设施的改善是很重要的，农村娱乐设施的建设提高了，直接影响着农民精神文明的提高，丰富了人民生活，可以提高居民的生活质量，还会提高农民的身体健康。然而从居民对所在村(镇)的娱乐设施的满意度调查结果来看(见图8-6)，居民对所在地娱乐设施的满意度评价较低，有很大一部分人不满意娱乐设施的建设，其中40.6%的人对所在地的娱乐设施不太满意，有16.0%的人对所在地的娱乐设施不满意，分别有7.8%和35.6%的人对所在地的娱乐设施感到非常满意和比较满意。

居民的居住环境包括住房面积、交通便利程度、环境绿化面积、商场数量、公园个数、基础设施建设等，所有以上方面都能体现居民居住环境的好坏，居住环境能直接反映居民生活现状，能更准确地说明新农

村建设以来居民的生活提高程度。从居民对现在的居住环境的满意度调查结果来看(见图 8-7),居民对居住环境的满意度评价较高,其中有 9.5%的人对所在地现在的居住环境非常满意,56.1%的人对所在地现在的居住环境比较满意。然而还有 25.7%的人对所在地现在的居住环境不太满意,有 8.7%的人对所在地现在的居住环境不满意,说明居民的居住环境还要进一步提高。

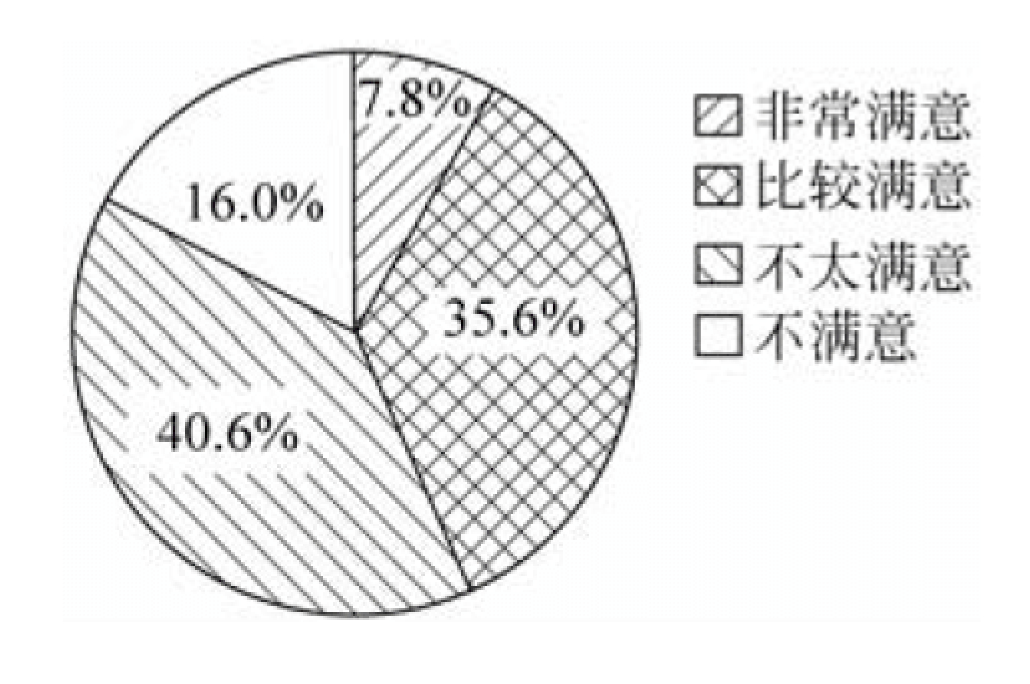


图 8-6　居民对所在村(镇)的娱乐设施的满意度评价

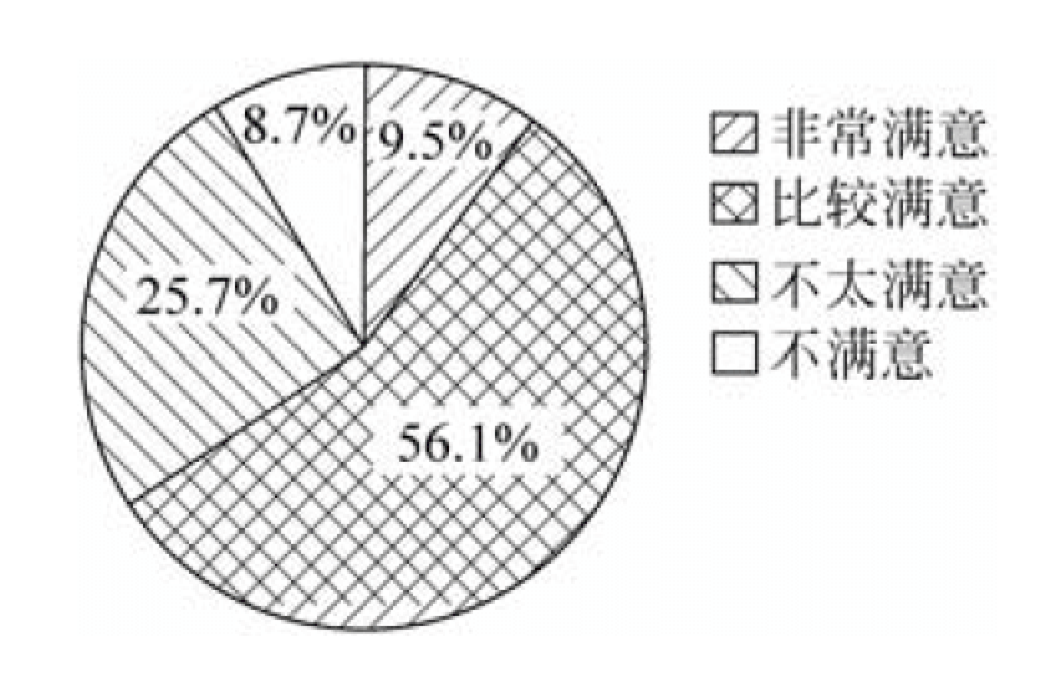


图 8-7　居民对现在的居住环境的满意度评价

8.2.4　农民教育文化水平的评价分析

本小节对新农村建设条件下农民的教育文化水平的现状进行分析,选择了两个方面进行描述性统计分析,分别是居民对所在村(镇)的教学条件的满意度评价和居民对目前所在村(镇)的教育环境的满意度评价,教学条件和教育环境最能体现目前农村教育现状,因此下面从这两个角度进行分析。

从居民对所在村(镇)的教学条件的满意度评价结果来看(见图 8-8),持满意态度和不满意态度的人数各占一半,其中对所在村(镇)的教学条件非常满意的人数占 6.3%,比较满意的人数占 46.0%,对所在村(镇)的教学条件不太满意的人数占 36.5%,对所在村(镇)的教学条件不满意的人数占 11.2%,说明新农村建设过程中农村教学条件还有待改善和提高。

从居民对目前农村的教育环境的满意度评价结果来看(见图 8-9),若把居民态度分成满意和不满意两种态度,则持满意态度的人数较多,其

中有 5.9%的人对目前农村教育环境非常满意，有 49.8%的人对目前农村教育环境持比较满意的态度，还有 37.2%的人对农村教育环境不太满意，有 7.1%的人对农村目前教育环境不满意。

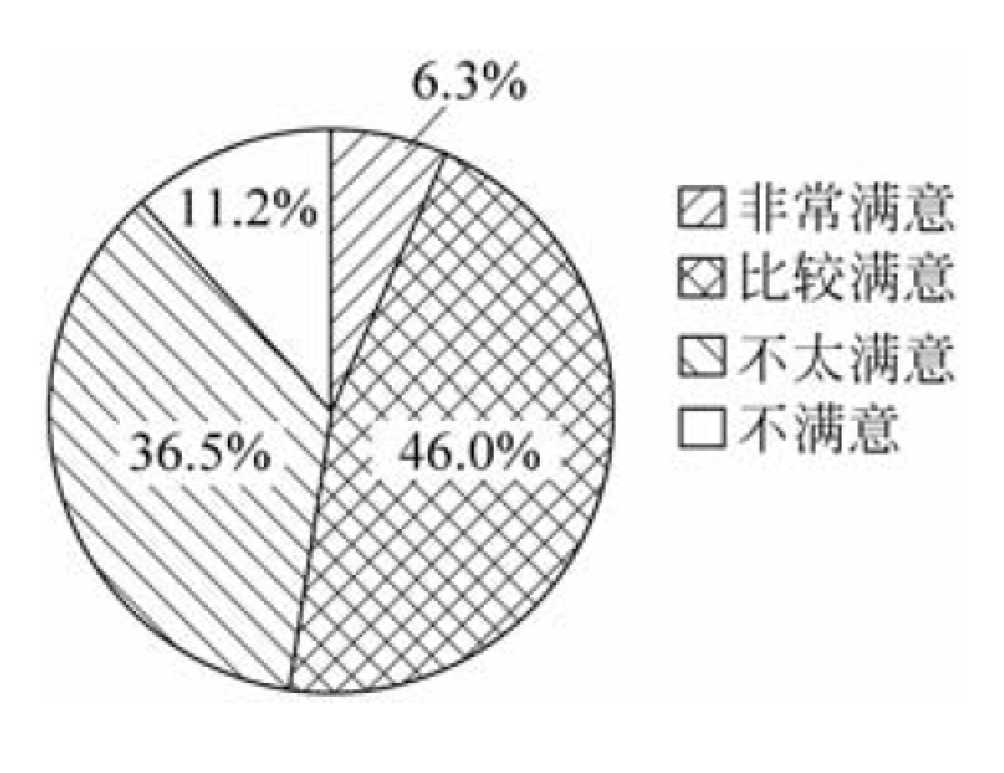


图 8-8　居民对所在村(镇)的教学条件的满意度评价

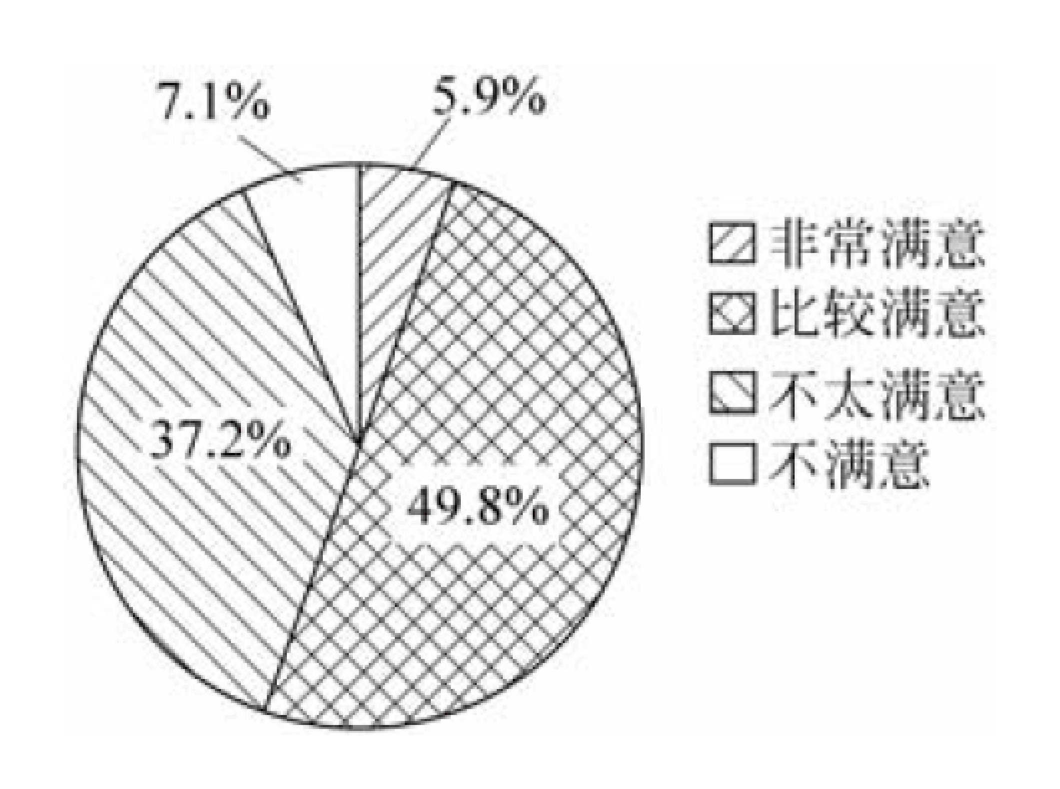


图 8-9　居民对目前农村的教育环境满意度评价

8.2.5　农民各项社会保障制度和社会治安的满意度评价

本小节对农民对各项社会保障制度和社会治安的满意度进行调查，从三个角度进行了简单的描述统计分析，第一是农民对所在村(镇)各种社会保障制度的满意度评价，第二是居民对所在村(镇)社会风气的调查，第三是居民对所在村(镇)社会治安状况的满意度评价。

首先，从居民对所在村(镇)各种社会保障制度的满意度评价结果来看(见图 8-10)，居民对社会保障制度的实施的评价较高，其中有 9.0%的人对目前农村社会保障制度的实施非常满意，54.5%的人对目前农村社会保障制度的实施比较满意，因此说居民总体上对目前农村社会保障制度的实施是满意的，但是仍然有 27.5%的人对目前农村社会保障制度的实施不太满意，9.0%的人对目前农村社会保障制度的实施不满意。

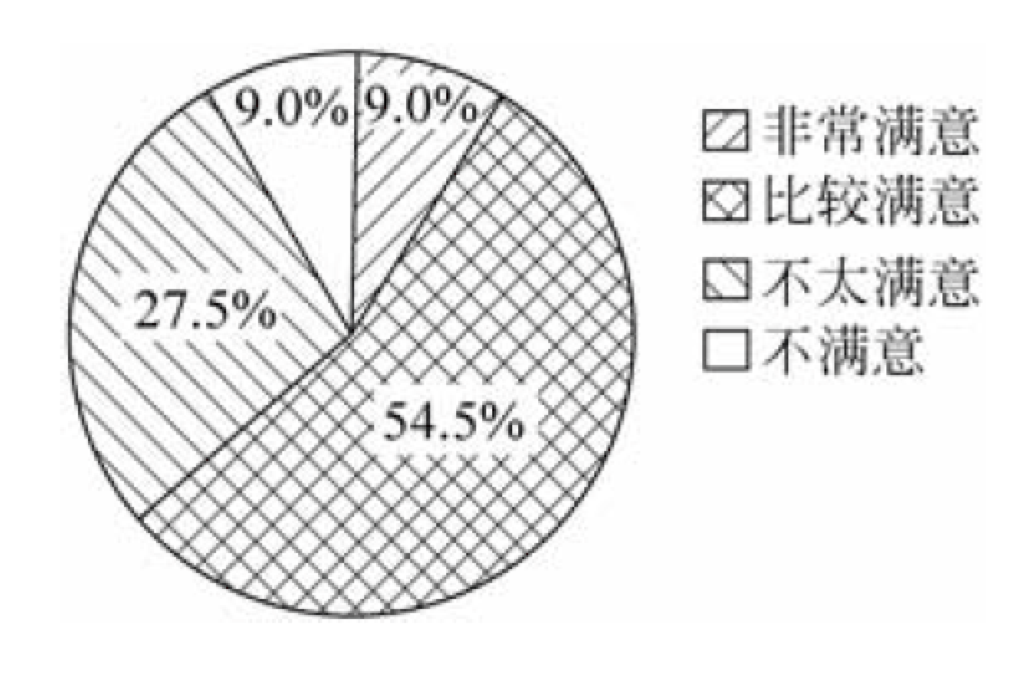


图 8-10　居民对所在村(镇)各种社会保障制度的满意度评价

其次从居民对所在村(镇)农村社会风气的调查结果来看,农村社会风气的好坏直接影响的是新农村建设的精神文明,也是新农村建设的一项主要内容。调查结果显示(见图 8-11),居民对新农村建设过程中目前农村社会风气的评价较高,其中 11.8%的人对所在村(镇)农村社会风气非常满意,62.0%的人对所在村(镇)农村社会风气比较满意,同时有少部分人对所在村(镇)农村社会风气评价偏低,其中 20.0%的人对所在村(镇)农村社会风气不太满意,有 6.2%的人对所在村(镇)农村社会风气不满意。

最后对居民所在村(镇)社会治安状况的满意度进行调查。社会治安是居民最关心的问题之一,都希望自己生活在安全的环境里。新农村建设以来,总体来说,居民对目前农村社会治安的评价较高(见图 8-12),10.4%的人对所在村(镇)社会治安状况非常满意,51.9%的人对所在村(镇)社会治安状况比较满意;仍然存在一部分人对所在地社会治安持不满意态度,其中有 25.4%的人对所在村(镇)社会治安状况不太满意,12.3%的人对所在村(镇)社会治安状况不满意。

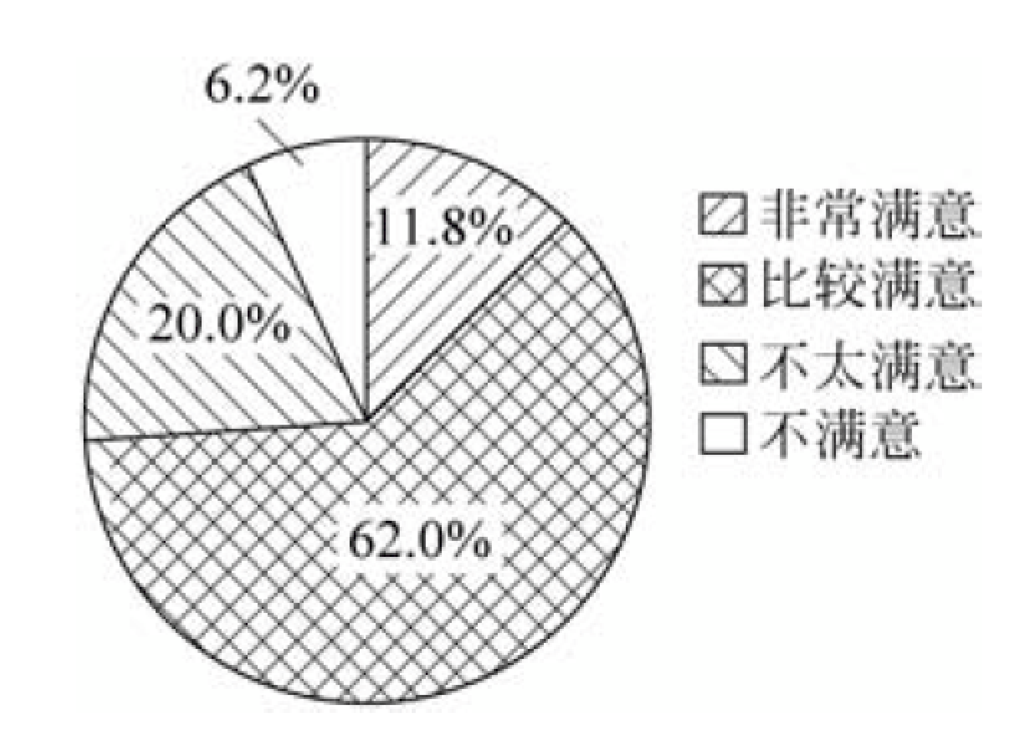


图 8-11　居民对所在村(镇)农村社会风气满意度评价

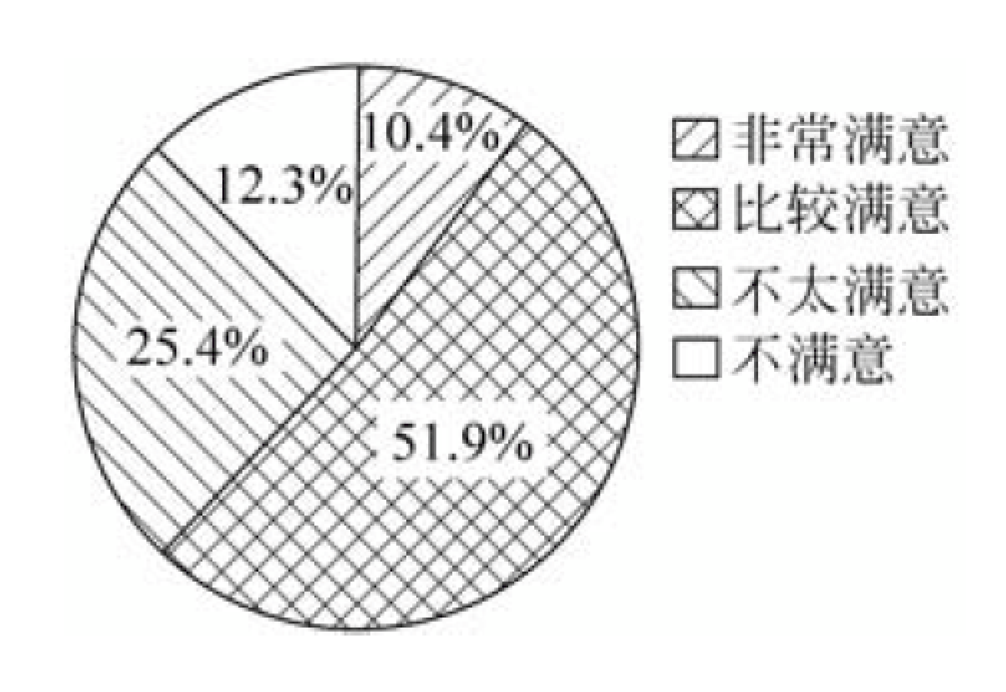


图 8-12　居民对所在村(镇)社会治安状况的满意度评价

8.2.6　村级班子的现状分析

农村干部班子的建设是新农村建设的一项内容,村级班子的办事能力直接体现在所管辖的农村的经济、生活、生产等各方面,干部们若为农民办实事,有经济头脑,那么所管辖的村镇一定会是发展很快的。

本研究从整体上对居民所在村(镇)班子干部的满意度进行调查,又因为大学生村官、选调生、公务员等的数量的增加,也推动着农村的发展,因此还对农民做了关于农民最希望大学生村官提供怎样的帮助的调查。

从居民对所在村(镇)班子干部的满意度调查结果来看(见图 8-13),有 13.0%的人对所在村(镇)班子干部非常满意,有 50.9%的人对所在村(镇)班子干部比较满意,这部分人对所在村(镇)班子干部评价较高;然而还有 23.4%的人对所在村(镇)班子干部不太满意,有 12.7%的人对所在村(镇)班子干部不满意。

很多大学生会选择回家乡当村官,为家乡做力所能及的贡献,促进家乡的经济发展,因此本次调查对农民做了相关调查,针对居民最希望大学生村官提供哪些方面的帮助做了调查。结果显示(见图 8-14),有 45.5%的人希望大学生村官给他们带来科技致富的知识,有 38.4%的人希望大学生村官给他们带来好的发展点子和思路,有 16.1%的人希望大学生村官能够配合村干部搞好村里的工作,新农村建设应该加强对大学生村官的培训,使他们尽快为农村建设做出贡献。

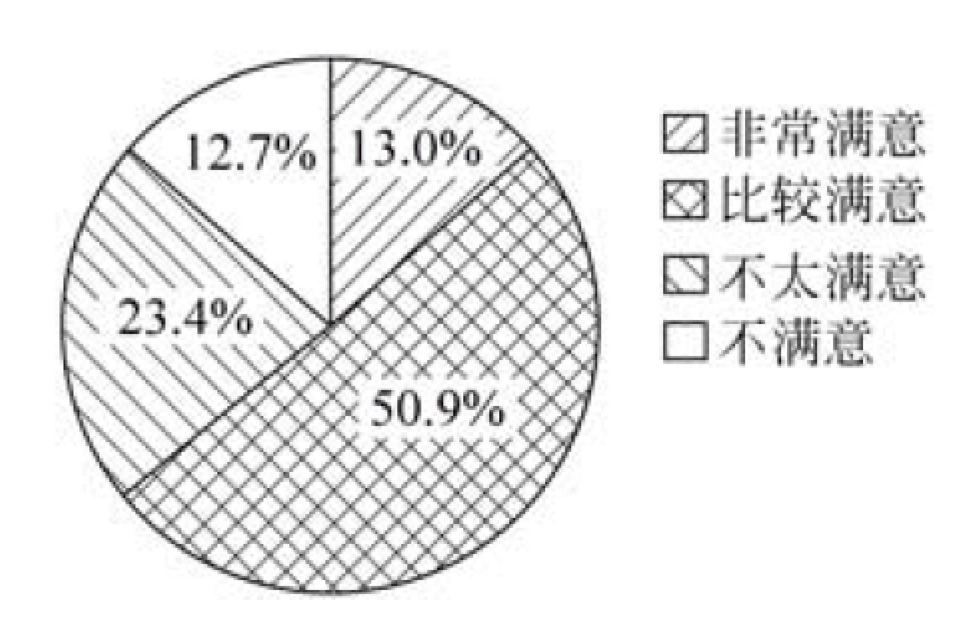


图 8-13 居民对所在村(镇)班子干部的满意度评价

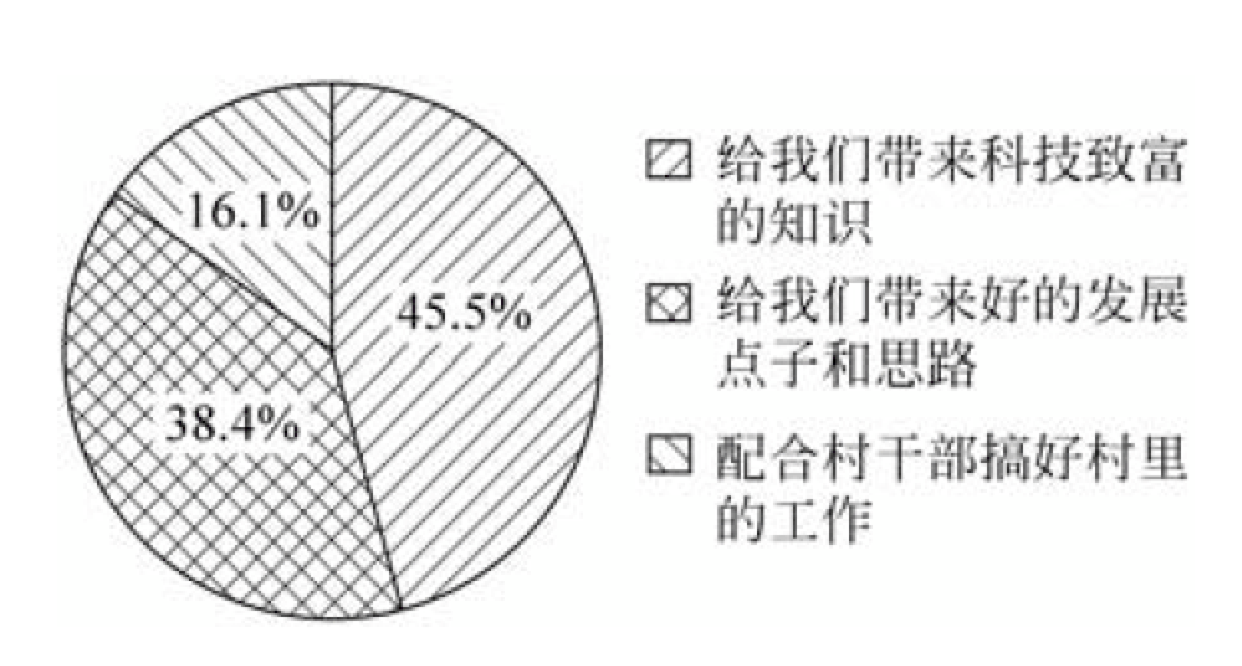


图 8-14 居民最希望大学生村官提供哪方面的帮助

8.3 居民对各项惠农政策措施的态度评价分析

国家对农民实行保护政策,近年来提出并实施了各项惠农政策,如工业反哺农业、城市支持农村政策,免除农业税政策,扶持畜牧业生产

补贴政策，提高农民素质培训工程，国家粮食综合补贴政策，免除中小学学杂费，一池三改工程等，新农村建设过程中更是大力实施国家政策，解决农民问题，切实提高农民的生活。因此本调查选取了前四个政策作描述性分析，分别探索居民对各项惠农政策的看法。

从居民对“工业反哺农业，城市支持农村”政策的看法的调查结果来看(见图 8-15)，认为“工业反哺农业，城市支持农村”政策正在逐步落实，但实施起来困难很大的人数占 40.1%，这部分人较多；有 22.4% 的人认为“工业反哺农业，城市支持农村”政策做得非常好，确实落到了实处；有 26.6% 的人认为“工业反哺农业，城市支持农村”政策基本没有落实；还有 10.9% 的人认为“工业反哺农业，城市支持农村”政策只停留在口头上。这说明新农村建设虽然取得了很大成就，但还有很多地方有待进一步提高。

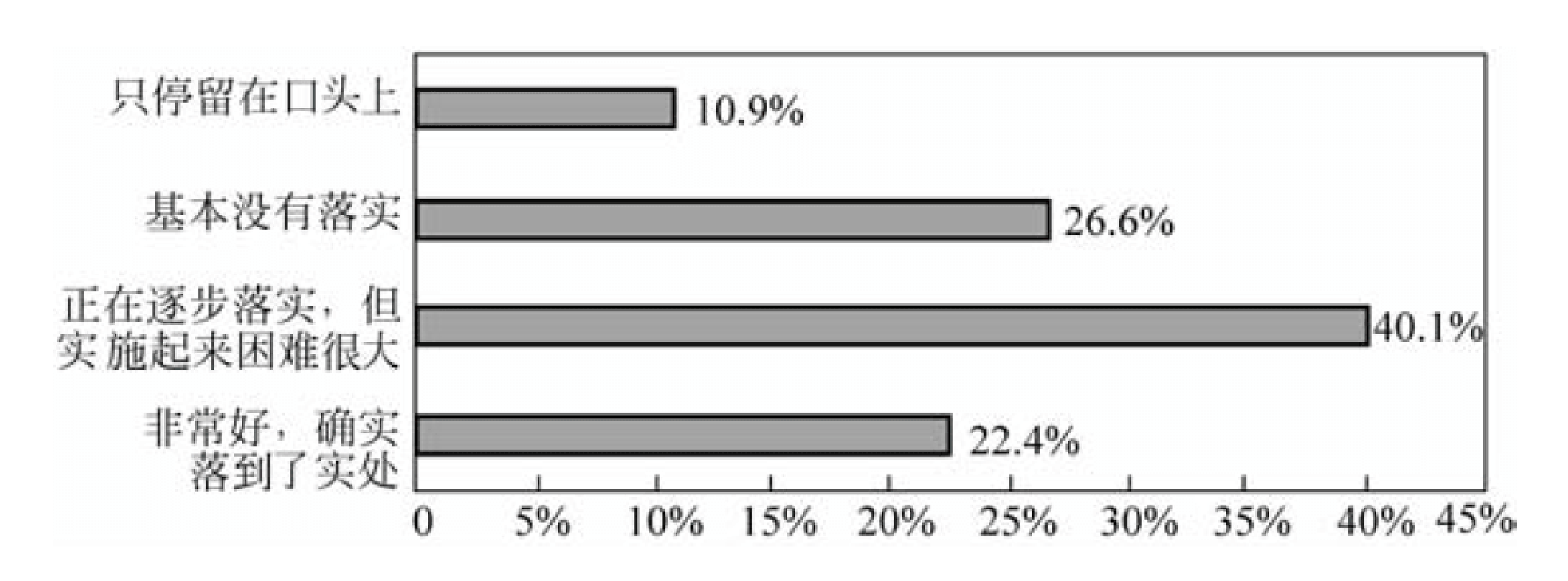


图 8-15　居民对“工业反哺农业，城市支持农村”政策的看法

从居民对免除农业税政策的看法的调查结果来看(见图 8-16)，农民对免除农业税政策评价非常高，认为免除农业税很好，农民种粮从此不用交税了的人数占比例 86.5%，仅有 13.5%的人对国家是否免除农业税无所谓。如此看来，新农村建设在支持农业发展方面的力度很大，免除农业税给农民带来了很多好处，相信农民未来生活会有很大的提高。

从居民对扶持畜牧业生产补贴政策的满意度评价结果来看(见图 8-17)，农民对国家扶持畜牧业生产补贴政策的评价较高，有 18.8% 的人对实施扶持畜牧业生产补贴政策非常满意，有 60.9% 的人对扶持畜牧业生产补贴政策的实施比较满意；然而还有一部分人对扶持畜牧

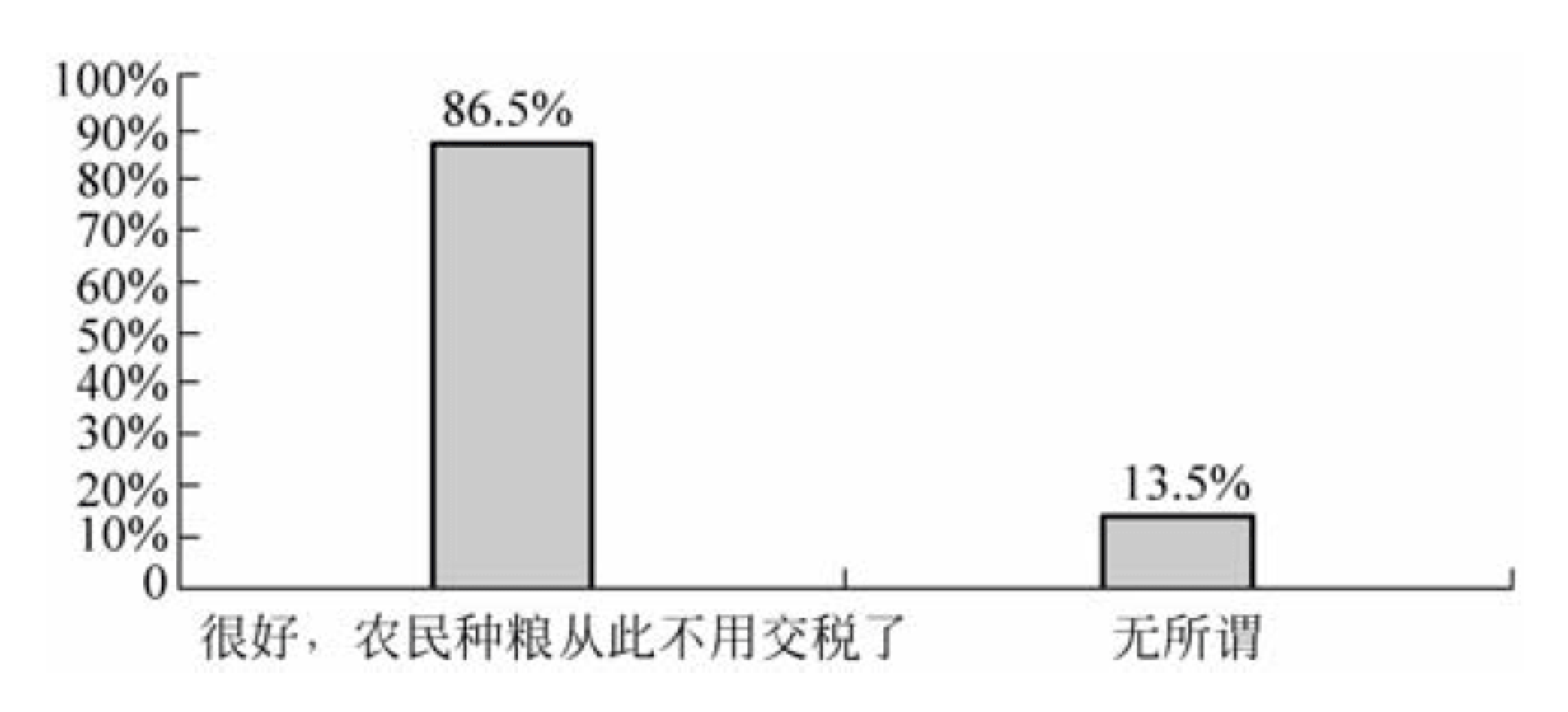


图 8-16 居民对免除农业税政策的看法

业生产补贴政策评价较低，其中有 16.5%的人对此项政策不太满意，有 3.8%的人对扶持畜牧业生产补贴政策不满意。

从居民对提高农民素质培训工程的满意度评价结果来看(见图 8-18)，有 14.2%的人对提高农民素质培训工程非常满意，有 54.4%的人对提高农民素质培训工程比较满意。总的来看，有很大一部分人对提高农民素质培训工程评价较高，新农村建设中把提高农民素质作为一项内容，农民自身的提高也是新农村建设的体现。有 26.4%的人对提高农民素质培训工程不太满意，有 5.0%的人对提高农民素质培训工程不满意。

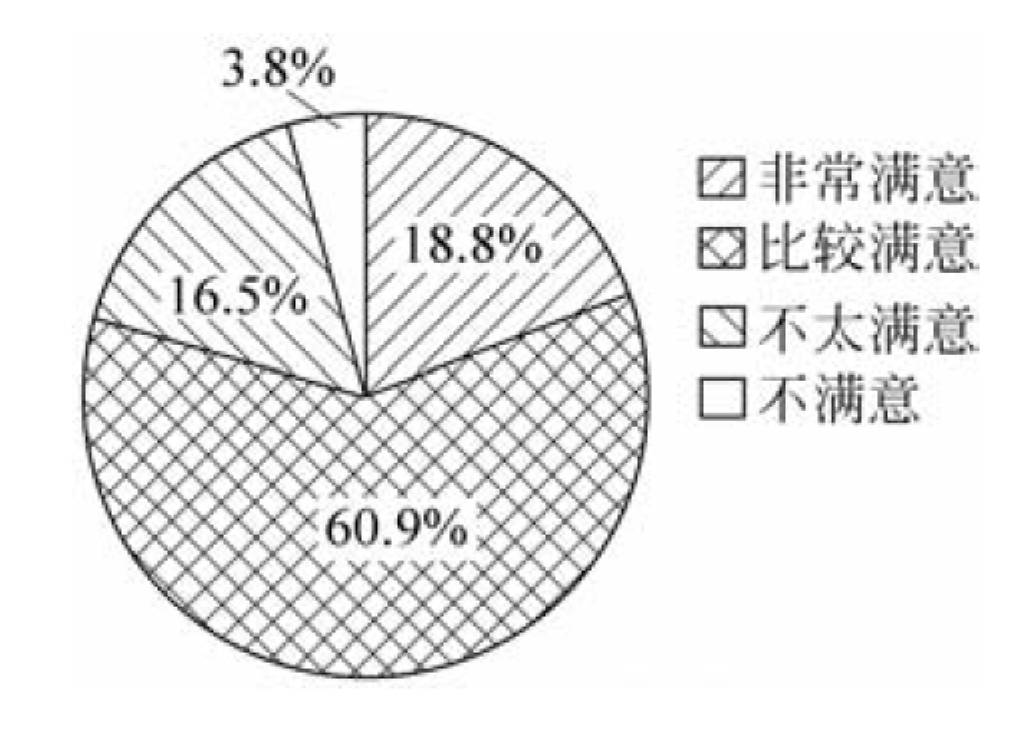


图 8-17 居民对扶持畜牧业生产补贴政策的满意度评价

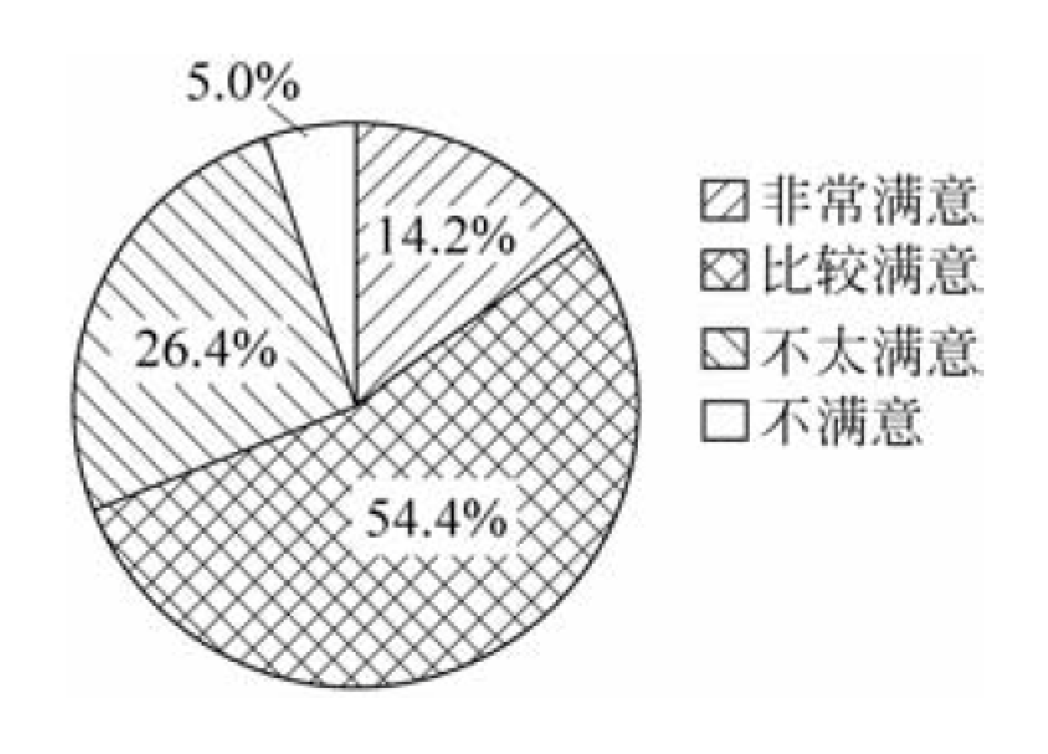


图 8-18 居民对提高农民素质培训工程的满意度评价

8.4 居民对所在村(镇)目前的新农村建设的看法

在对居民关于所在村(镇)目前的新农村建设的调研中,认为所在村(镇)做到了某项要求的居民选择“是”,认为没有做到某项要求的居民选择“否”。

本次针对八项内容做了调查,调查结果见表 8-5。从结果中来看,认为所在村(镇)目前的新农村建设受到领导重视,并且投入到位的人数占 58.5%;认为所在村(镇)目前的新农村建设因地制宜、措施得力的人数占 60.2%;认为所在村(镇)目前的新农村建设获得群众拥护,信心十足的人数占 60.4%;认为所在村(镇)目前的新农村建设真抓实干、发展健康的人数占 58.5%;认为所在村(镇)目前的新农村建设急于求成、事与愿违的人数占 35.9%;认为所在村(镇)目前的新农村建设注重形式、收效甚微的人数占 41.0%;认为所在村(镇)目前的新农村建设困难重重、信心不足的人数占 40.3%;认为所在村(镇)目前的新农村建设什么也没干的人数占 27.8%。

表 8-5 居民对所在村(镇)目前的新农村建设的看法 %

项目	领导重视、投入到位	因地制宜、措施得力	群众拥护、信心十足	真抓实干、发展健康	急于求成、事与愿违	注重形式、收效甚微	困难重重、信心不足	好像什么也没有干
是	58.5	60.2	60.4	58.5	35.9	41.0	40.3	27.8
否	41.5	39.8	39.6	41.5	64.1	59.0	59.7	72.2

总的来看,居民对目前新农村建设取得的成就是予以肯定的,新农村建设给农民带来了很多好处,各个村(镇)也大都切实把新农村建设作为重大项目进行,认认真真干实事,虽然还存在着很多不足,但是随着时代的进步,相信新农村建设会取得更大的成绩。

8.5 新农村建设面临的主要困难评价分析

本节针对新农村建设过程中面临的主要困难进行分析,主要从九大方面来探讨主要困难问题,分别是缺乏资金;农民致富难;农民就业

难；发展现代农业难；发展农村二、三产业难；壮大村级集体建设难；农民素质提高难；构建“工业反哺农业，城市扶持农村”长效机制难；缩小城乡差距难。在每一个大的方面里分别列举几个重要原因，进行简单的描述分析，具体的汇总结果见表8-6。

表8-6　新农村建设面临的主要困难

方　面	原　因	频率/%
缺乏资金	集体经济薄弱，村里没有钱	40.6
	政府资金投入少	44.6
	老百姓没钱	35.6
农民致富难	缺乏资金	62.2
	缺少技术	57.3
	信息不灵	32.6
农民就业难	没有文化，很难找到合适的工作	50.7
	本地经济发展落后，不能提供充分的就业岗位	48.4
发展现代农业难	耕地少，地力差	29.9
	小规模分散经营	34.2
	农业适用技术推广难	28.8
	农产品市场需求信息不灵	28
发展农村二、三产业难	无技术	67.5
	无人才	53.1
	无资金	43.2
壮大村级集体建设难	原来就没有基础，现在无从下手	30.7
	市场竞争激烈，要想壮大，不容易	31.8
	缺乏资金和人才	61.6
农民素质提高难	农民认识不到知识的重要性	34
	职业教育发展滞后	34.2
	农民可以接受培训的机会太少了	68.4
构建“工业反哺农业，城市扶持农村”长效机制难	这是一项大工程，需要各方面的努力	47.7
	政府在这方面的工作力度还不大	40.5
缩小城乡差距难	城市各方面条件优越，经济发展越来越快，农村反之	53
	政府在加快农村发展方面的工作力度不够，措施还不多	39.4

结果显示，有 44.6%的人认为资金缺乏的主要原因是政府资金投入少，有 62.2% 的人认为农民致富难的主要原因是缺乏资金，有 43.2%的人认为发展农村二、三产业难的主要原因是无资金，有 61.6%的人认为壮大村级集体建设难的主要原因是资金缺乏和无人才，从这几方面综合来看，资金的投入对新农村建设的发展有着很大的影响。

有 57.3%的人认为农民致富难的另一重要原因是缺少技术，有 28.8%的人认为发展现代农业难的一个重要原因是农业适用技术推广难，有 67.5%的人认为发展农村二、三产业难的一个重要原因是无技术，综合几方面来看，技术也是推动新农村建设发展的重要动力。

有 32.6%的人认为信息不灵同样是农民致富难的一个重要原因，有 28%的人认为农产品市场需求信息不灵是发展现代农业难的一个重要原因，综合两个方面来讲，获得充分的信息是新农村建设发展的重要保证。

8.6　居民对新农村建设的企盼分析

新农村建设以来，不仅取得了显著的成效，也使得广大人民群众对新农村建设越来越有信心，对它的期望也越来越大，并且每个人的企盼都大不相同。本次调查针对居民对新农村建设的企盼做了研究分析，并且对居民认为如何开始着手开展新农村建设以及居民早日致富的渠道做了调查，下面分别对调查结果进行简单的描述性统计分析。

如何建设好新农村建设，该从哪方面着手，是国家和人民非常关心的问题。调查结果显示（见表 8-7），认为新农村建设最应该首先从促进农民增收的人数占比例 33.0%，有 28.1%的人认为第二个重要的应该是健全农村医疗卫生、养老、最低生活等各项保障制度，提高农民素质和加强基础建设。整治村容村貌同样是新农村建设应该开展的工程，有 24.1%的人认为开展新农村建设要有个好的带头人，只有好的领导才能起到大作用。

表 8-7　村(镇)开展新农村将建设应如何入手

项　　目	频率/%
推进土地流转,实现规模经营	16.1
发展现代农业	19.6
发展农村工业	20.3
促进农民增收	33.0
加强基础建设,整治村容村貌	23.6
因地制宜,搞好农村规划	20.5
加强村级组织建设	14.6
提高农民素质	24.0
加强村级文化建设	14.4
移民搬迁	5.7
健全农村医疗卫生、养老、最低生活等各项保障制度	28.1
加强对新农村建设工作的领导	11.3
因地制宜,循序渐进	9.0
要选一个好的带头人	24.1
其他	3.1

仅有 5.7%的人认为新农村建设过程中应该把移民搬迁作为着手处理的工作,说明移民搬迁工作应该放在次要地位;仅有 9.0%的人认为进行新农村建设应该因地制宜、循序渐进,这个比重很低,但是作者认为因地制宜、循序渐进是正确的道路,新农村建设不应急功近利,要高效率地进行。

从居民对新农村建设的企盼的调查结果来看(见表 8-8),居民对新农村建设有很大的企盼,有 44.6%的人对加快小城镇建设;发展二、三产业;提供就业就会存在期望,这是农民对新时代新农村建设的首要企盼,也是与农民生活息息相关、紧密相连的主要内容;其次有 39.6%的人希望尽快壮大农村集体经济、发展公益事业、改善生存环境,这是新农村建设过程中农民第二个想要的企盼;在新农村建设过程中居民对构建“工业反哺农业,城市支持农村”机制落在实处报的企盼值相对来说较低,仅有 23.4%的人对其抱有期望。总体来说,农民对新农村建设的期望值较高,充满信心。

表 8-8　居民对新农村建设的企盼

项　　目	频率/%
现有的“三农”政策稳定不变，未来的政策更加完善	32.5
增加农业投入，搞好兴水沃土建设，提高农业综合生产能力	34.2
加快小城镇建设，发展二、三产业，提供就业机会	44.6
壮大集体经济，发展公益事业，改善生存环境	39.6
增加社会事业投入，构建社会保障体系	32.6
构建“工业反哺农业，城市支持农村”机制落在实处	23.4
加强对新农村建设工作的领导	26.9

从居民关于通过什么渠道早日致富的调查结果来看（见表 8-9），有 64.8%的人认为最主要的是通过获得更多技术和信息支持来早日致富，新时代新农村建设把提高农民技术水平和获得信息的广泛渠道作为一项重要工程，所以技术是农民发家致富的重要手段；其次有 50.9%的人把获得稳定的就业渠道作为发家致富的第二个重要手段；而居民把获得平等的国民待遇作为致富渠道的人很少，可见平等的国民待遇并不能像更多的技术和信息支持一样能够快速获得财富。新农村建设的渠道很多，合理利用将会为农民带来更多的利处。

表 8-9　居民希望早日致富的渠道

项　　目	频率/%
通过有偿流转获得更多的土地或养殖更多的牲畜，实现规模经营	27.3
获得更多的技术或信息支持	64.8
获得稳定的农字号龙头企业或专业协会的带动保障	29.2
获得稳定的就业渠道	50.9
通过参加各类培训提升自己的综合素质	39.2
获得平等的国民待遇	18.8

第 9 章　基于决策树的新农村建设满意度细分模型分析

9.1　决策树简介

决策树是具有分类功能的一种技术，并根据分类结果对样本记录进行预测，它提供了一种展示类似在什么条件下会得到什么值这类规则的方法。决策树算法最早是由 Hunt 等人提出的 CLS 算法，接着又诞生了 ID3、CART、C4.5、C5.0、ID4、SPRINT、QUEST 等多种决策树算法。在实际应用中，决策树技术主要用于分类和预测。

CART(classification and regression trees)也称为分类回归树，由于其每个非叶子节点都有两个子节点，是一种二叉决策树的模型。CART 算法用在目标变量是分类变量时，它就是分类树；如果目标变量是范围变量，则为回归树。在本研究中用到的 CART 在选择分割属性时以 Gini 系数作为选择标准。

设 M 为一数据样本集，且 M 中所含的样本数记为 C，A 为 M 中的一个类别，P_A 为类 A 在 M 中出现的频率。那么对于样本集 M 的 Gini 系数就被定义为

$$\mathrm{Gini}(M)=1-\sum P_A^2 \tag{9-1}$$

其中，Gini(M)越小，分支所产生的子集纯度就越高。

如果用属性 B 将 M 划分为两个子集 M_1 和 M_2，其中 C_1 为集合 M_1 中的样本个数，C_2 为集合 M_2 中的样本个数。那么，基于这个划分的 Gini 系数为

$$\mathrm{Gini}(M)=\frac{C_1}{C}\times\mathrm{Gini}(M_1)+\frac{C_2}{C}\times\mathrm{Gini}(M_2) \tag{9-2}$$

1. 构建树

CART 算法的基本迭代过程是：创建一个节点，CART 的训练将从此节点开始，并把此节点称为根节点 A；设 S 为训练样本，它是包含 s 个数据样本的集合。如果 S 中的所有 s 都在同一类中，则算法停止，把根节点 A 改成树的叶节点，并用该类标记，否则，从众多的属性变量中选择一个当前最佳的分支变量，此变量必须满足使异质性下降最快，对分支异质性的度量，采用 Gini 系数；将训练样本 S 划分成两个子集 S_1 和 S_2，此时 S_1 和 S_2 中的数据比父节点的数据更同质，即差异性更小；对上述过程 CART 算法使用同样的过程递归，S_1 和 S_2 将被继续分别划分成两个不同的子集。相同的预测变量在树的各层中可以不断地被使用。递归过程不断重复，直到达到同质性标准或触发终止条件，CART 的生长将停止。

CART 算法的终止条件为：①节点达到完全纯性，即节点中样本的目标变量取值差异性很小或是不存在；②树的深度达到用户指定的深度，即树的层数达到了用户指定的层数；③节点中样本的个数少于用户指定的个数。有时，虽然节点的纯度未达到完全纯性，树的深度也没有达到用户指定的要求，但是节点中样本的个数已经非常少了，此时再继续划分下去已经没有意义。用户可以通过指定节点样本的个数来终止划分，即停止树的生长。

2. 分类回归树的修剪

分类树完全生长过程中会出现过度拟合现象。需要通过剪掉树叶节点，使模型能够对新数据进行更准确有效的分类。本研究采用 CART 系统的成本-复杂度最小原则（minimal cost-complexity pruning）对分类回归树进行后剪枝，最后将会构建一个节点数目递减的有限有序子树序列。其测度表示为

$$Ra(M) = Ra(M) + a \times |M| \tag{9-3}$$

其中，$|M|$ 为该树的叶节点的个数，a 为复杂度参数，$Ra(M)$ 理解为该树加权错分率与对复杂度处罚值之和的复合成本。

3. 分类回归树评估

一般评估方法主要为采用建立评估图节点和通过对验证集进行对比来测试模型的优劣。

9.2　基于决策树的新农村建设满意度细分模型的建模过程

1. 问题理解

在对新农村调研问卷数据进行数据挖掘之前，首先要对挖掘目标有一个充分的理解。本研究的此部分主要目的在于用决策树模型挖掘出的规则以发现农村居民中的满意者和非满意者。本研究中农村居民满意程度群细分是在借鉴了市场调查研究中著名的市场细分理论基础上进行的。市场细分的概念是美国市场学家温德尔·史密斯于 20 世纪 50 年代中期提出的，指按照消费者欲望与需求把一个总体市场划分成若干个具有共同特征的子市场的过程。本章将细分变量分为广度细分变量和深度细分变量。其中广度细分变量就是参与调研农村居民的基本信息，通过广度细分变量新农村居民满意度人群划分可以直接形象地得到描述。而深度细分变量是参与调研农村居民对影响总体满意程度的生产、居住环境、村庄管理等较深层次问题的认可状况，是具体按照“二十字方针”的要求确定的四个变量。决策树模型可以将人们不能通过观察直接得出的信息从数据中挖掘出来，形成规则，利用这些规则寻找到可能的农村满意人群和不满意人群。文中以问题“总体来说是否满意”(变量名为 C_{13}，问卷中为十题第 1 问)为目标变量，广度细分变量和深度细分变量设置见表 9-1 和表 9-2。

表 9-1　广度细分变量

变量名	问卷内容(题号)	对应问卷选项
A_1	性别(一)	(1)男；(2)女
A_2	年龄(二)	(1)30 以下；(2)30～45；(3)45～60；(4)60 以上

续表

变量名	问卷内容(题号)	对应问卷选项
A_3	文化程度(三)	(1)不识字或少识字;(2)小学;(3)初中;(4)高中;(5)大专及以上
A_4	职业(四)	(1)农民群众;(2)村干部;(3)乡镇干部;(4)其他
A_5	政治面貌(五)	(1)党员;(2)民主党派;(3)其他
A_6	所在村类型(六)	(1)示范村;(2)达标村;(3)移民村;(4)其他村
A_7	居住类型(七)	(1)租房;(2)土坯;(3)砖瓦;(4)小楼;(5)居民区

表 9-2　深度细分变量

变量名	问卷内容及题号	对应问卷选项
C_5	收入(十题第 4 问)	(1)非常满意;(2)比较满意;(3)不太满意;(4)不满意
C_{11}	居住环境(十题第 5 问)	(1)非常满意;(2)比较满意;(3)不太满意;(4)不满意
C_{15}	社会风气(十题第 14 问)	(1)非常满意;(2)比较满意;(3)不太满意;(4)不满意
C_{17}	领导干部(十题第 12 问)	(1)非常满意;(2)比较满意;(3)不太满意;(4)不满意

2. 定义数据源

在进行数据预处理后,本例中共收集到有效问卷数量为 905 份。这 905 条记录组成的样本集所对应的属性共有 11 个,基本信息属性 7 个,细分满意度属性 4 个。目标变量设为居民总体满意度的变量。利用文中数据预处理部分的变量选择节点和过滤节点,得到处理之后的数据集"决策数据. xls",并将此数据定义为数据源。

3. 建立模型

由于模型在建立好之后,需要用一定量的数据对它进行检验,经过实际数据检验并被证明正确的模型应用到实践中才有意义,因此,在数据分析阶段,需要将新农村问卷调研数据集分为训练集和验证集两个

部分，其中训练集用来建立模型，验证集用来对所建模型进行评估。加入 Sample 节点作随机抽样，固定随机数种子（本例中为 177 216），由于样本量不是很大，为了使模型更趋于准确，故从数据源中抽取 85% 的数据作为训练集，仅使用剩下的 15% 作为验证集。本研究采用 CART 决策算法，应用 Clementi 12.0 建立模型并进行分析，具体流程图见图 9-1。

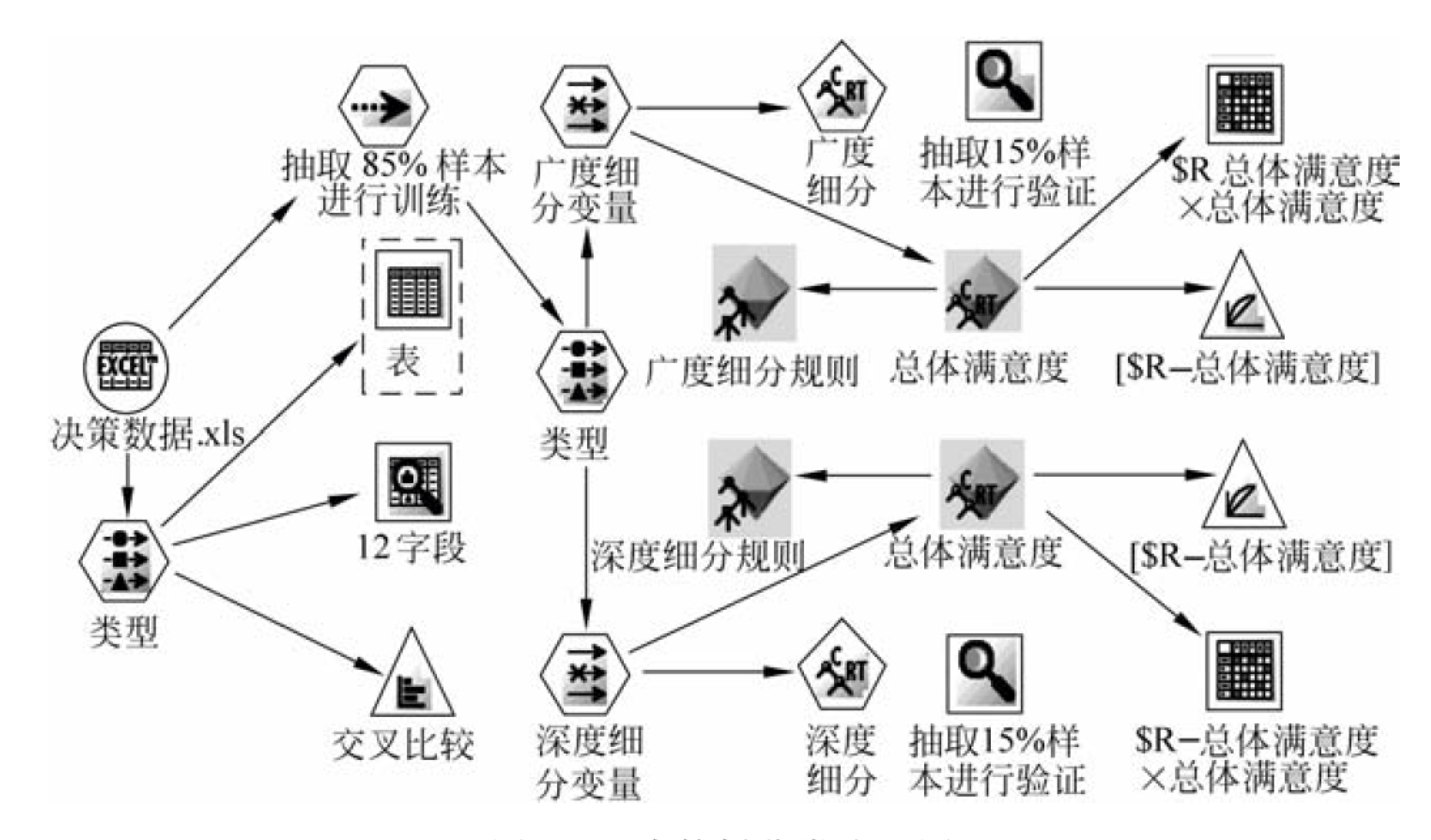


图 9-1　决策树分类流程图

4. 调研数据的基本信息

从预处理之后的数据来看（见表 9-3），参与运算的广度细分变量中，男性为 465 人，占总人数的 51.4%，女性为 440 人，占总人数的 48.6%，参与调查者男女比例大致相等；从年龄来看，30 岁（不含 30 岁）以下有 291 人，30 至 45 岁（不含 45 岁）有 275 人，45 至 60 岁（不含 60 岁）有 241 人，60 岁以上有 98 人；从文化程度来看，不识字或者少识字者 63 人，小学文化程度 149 人，初中文化程度 365 人，高中文化程度 290 人，大专及以上文化程度 38 人；从职业来看，农民占了大多数，达到 619 人，剩余为村干部、乡镇干部和部分外来务工者等；从政治面貌来看，参与调查的群众为 757 人，党员为 116 人，民主党派为 32 人；另外，所有参与调查的人员中，有 441 人来自新农村示范村，占

总人数的 48.8%，这与制定抽样计划人数大致相符，其余 51.2%来自新农村达标村、移民村和其他村；居住类型以砖瓦房居民居多，达到总人数的 38.4%，28.6%的居住类型为居民区，20.8%为自建小楼，其余为租房和土坯房。详细信息见表 9-3。

表 9-3　广度细分变量情况表

变量名	变量内容	细　　分	人数	比例/%
A_1	性别	男	465	51.4
		女	440	48.6
A_2	年龄	30 以下	291	32.1
		30～45	275	30.4
		45～60	241	26.6
		60 以上	98	10.9
A_3	文化程度	不识字或少识字	63	7.0
		小学	149	16.5
		初中	365	40.3
		高中	290	32.1
		大专及以上	38	4.1
A_4	职业	农民	619	68.4
		村干部	32	3.1
		乡镇干部	31	3.5
		其他	223	25.0
A_5	政治面貌	党员	116	12.7
		民主党派	32	3.6
		其他	757	83.7
A_6	所在村类型	示范村	441	48.8
		达标村	55	6.1
		移民村	173	19.1
		其他村	235	26.0
A_7	居住类型	租房	68	7.5
		土坯	42	4.6
		砖瓦	348	38.4
		自建小楼	188	20.8
		居民区	259	28.6

参与分析的905份问卷中，对当前总体状况满意度为“非常满意”的农村居民有53人，占总数的5.9%；持“比较满意”态度的农村居民有453人，占总数的49.9%；持“不太满意”态度的农村居民有336人，占总数的37.1%；持“不满意”态度的农村居民有63人，占总数的6.9%。我们将这四个值记为农村居民满意程度的平均水平。另外，影响总体满意度的4个细分满意度属性情况见表9-4。

表9-4　影响总体满意度因素情况表　　%

变量名	变量内容	选项			
		(1)非常满意	(2)比较满意	(3)不太满意	(4)不满意
C_5	收入	9.0	54.5	27.4	9.1
C_{11}	居住环境	11.8	61.9	19.9	6.4
C_{15}	社会风气	9.5	56.1	25.7	8.7
C_{17}	领导干部	9.5	56.3	27.3	6.9

5. 模型设定

在利用CART算法建立决策树的过程中，首先使用CART简单模式对数据进行探索性分类(具体结果不再赘述)，第二步使用专家模式，具体设定如下：树的最大深度设为5(即根节点下的最大层数为5)，最大代用项为5；设定分类目标的杂质项测量系数为Gini系数，并指定在树的生长过程中进行新的分割所需的最小杂质改变量为0.0003，若某一分支的最佳分割所引起的杂质改变量低于0.0003，则决策算法程序不会继续进行分割；选择修建树选项，即当进行一个节点的分支的构建时，不能显著提高树的准确性，则自动剔除此节点以下的所有分支；使用百分比的方法来控制树的生长，设定父节点记录数占总记录数为最少2%，子节点记录数占总记录数为最少1%。在此设定下，得到最终决策树结果。

6. 模型结果

(1) 新农村建设满意度的广度细分模型

广度细分输出结果见图9-2和图9-3。

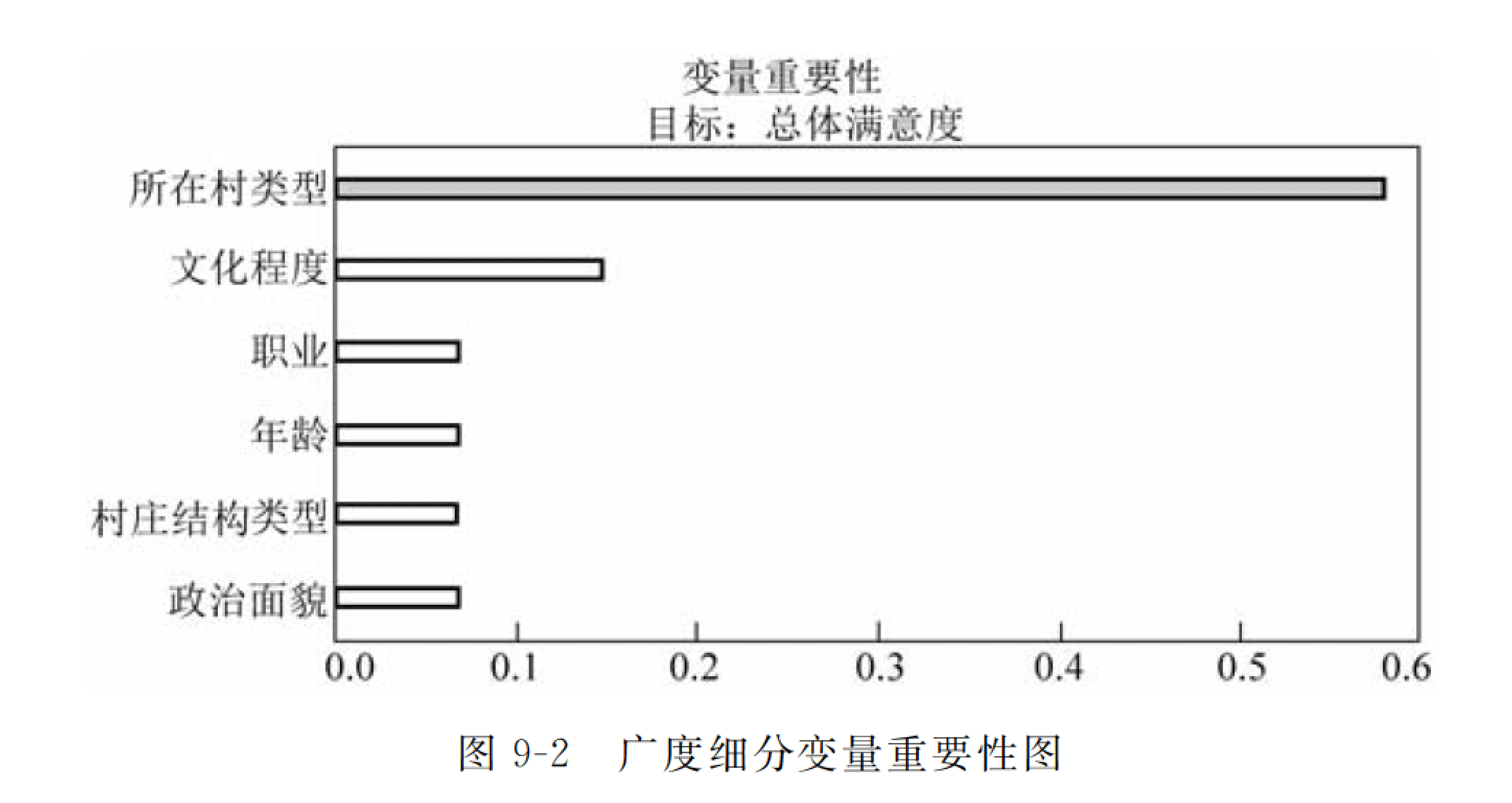


图 9-2　广度细分变量重要性图

由图 9-2 中可以看出，所在村庄类型为最主要变量，文化程度为次要变量，除了“性别”变量基本无影响之外，其余人口统计学变量影响程度相当。

由广度细分决策树状图 9-3 可以将农村居民分为以下两类：

① 参与调查的农村居民中持肯定态度的人群。我们将选择“非常满意”和“比较满意”的农村居民的比例高于平均比例的划分人群定义为对当前生活状态持肯定态度的人群。由图 9-3 可以发现，节点 1 代表的来自新农村(新民居)示范村的农村居民信息，持“非常满意”的人数为 37 人，所占此群体比例为 8.981%，高出参与调查总体农村居民 3.08 个百分点；持“比较满意”的人数为 227 人，所占此群体比例为 55.097%，高出参与调查总体农村居民 5.197 个百分点[①]。值得注意的一点是，在所有参与调查的农村居民中，选择“非常满意”的有 44 人，在节点 1 中的显示结果说明在新农村(新民居)示范村中有37 人选择了“非常满意”，占参与运算数据全部选择此项的农村居民人数的

① 在此部分及以下的分析中，均将各个节点中比例和总体比例相比较。而不是与随机抽取的 85%(即节点 0)数据得出的比例相比较。原因有以下两点：一是通过将子节点比例和节点 0 与样本总体比例相比较，数值虽然有差异，但是差异不大，而总体样本数据更能说明问题，能更好地代表均值；二是在文中模型验证部分(根据剩余 15%样本)得出了模型的准确性，准确率相对较高。

84.0%。

② 参与调查的农村居民中持否定态度的人群。我们将选择“非常满意”和“比较满意”农村居民的比例低于平均比例的划分人群定义为对当前生活状态持否定态度的人群。由图 9-3 可以发现,节点 2 代表的是来自新农村达标村、移民搬迁村和其他类型村庄的农村居民信息,在此类人群中,持“非常满意”态度的人数为 7 人,所占此群体比例为 1.818%,低于参与调查总体农村居民平均水平 4.082 个百分点;持“比较满意”态度的人数为 171 人,所占此群体比例为 44.416%,低于参与调查总体农村居民平均水平 5.484 个百分点。在节点 5 和节点 6 所反映的情况是,自身文化程度为小学、初中的农村居民中,“非常满意”态度的人数为 7 人,所占此群体比例为 3.097%,持“比较满意”态度的人数为 110 人,所占此群体比例为 48.673%,其满意度略高于自身文化程度为不识字或者少识字、高中或者大专以上的农村居民,分别高出了 3.097%和 10.308%。我们可以称节点 5 所代表人群为文化程度趋中人群,节点 6 代表文化程度相对低和相对高的人群。结果显示,非新农村(新民居)示范村居民满意程度均低于新农村(新民居)示范村居民,非新农村(新民居)示范村居民中,文化程度趋中的人群满意程度高于文化程度相对低和相对高的人群。

(2) 新农村建设满意度的深度细分模型

深度细分输出结果见图 9-4 和图 9-5。

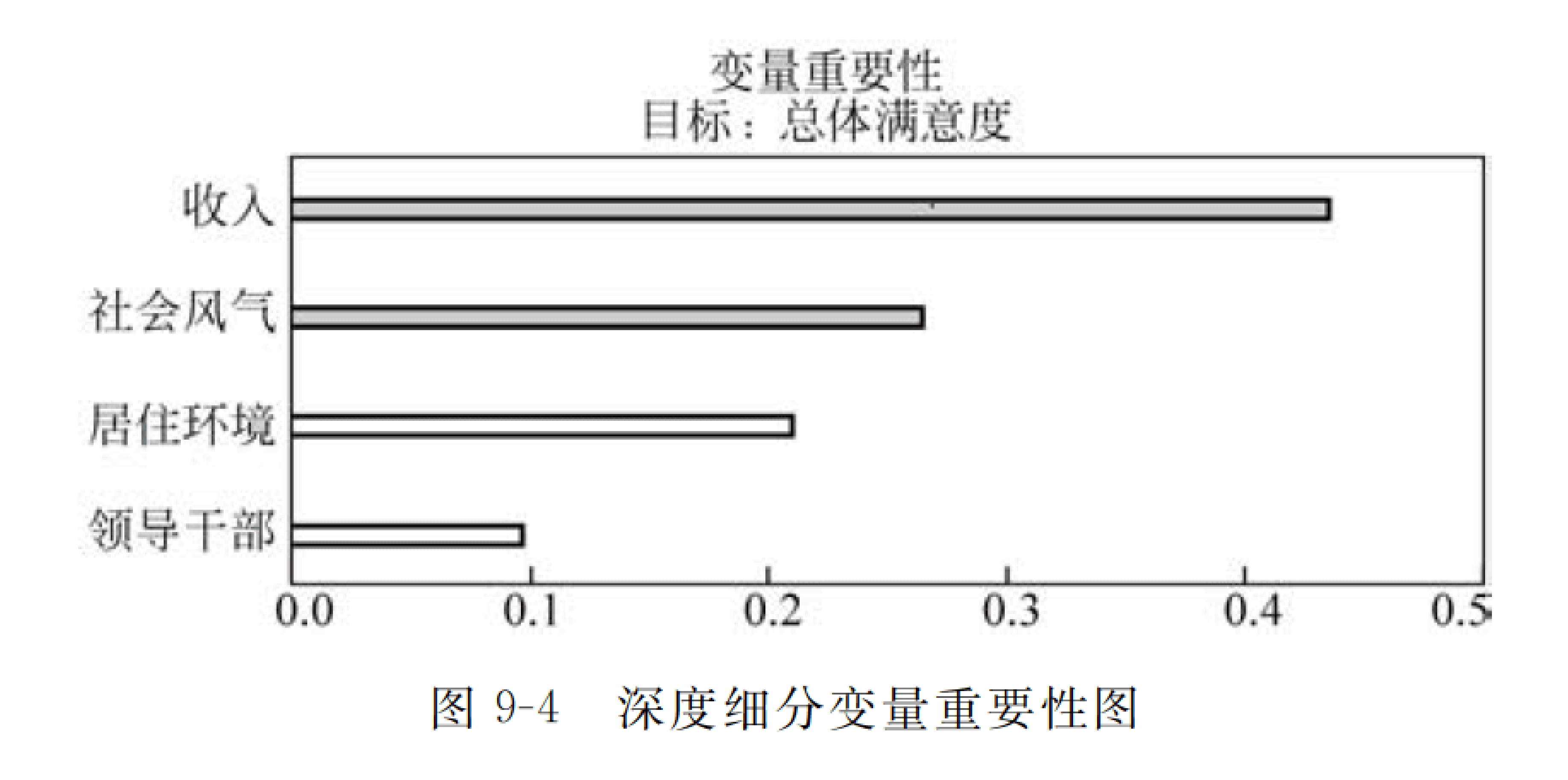


图 9-4　深度细分变量重要性图

由变量重要程度图中可以看出,农民收入为最主要变量,社会风气

和居住环境为次要变量，领导干部的影响程度略小于其他变量。

由深度细分决策树状图 9-5 可以得出：第一层细分变量为农村居民收入满意程度。收入因素将参与训练的样本分成对比鲜明的两个群体。对比决策树中收入因素产生的节点 1 和节点 2 发现，节点 1 中对收入满意程度问项选择“非常满意”、“比较满意”甚至“不太满意”的农村居民，在总体满意程度问项中选择“非常满意”和“比较满意”的农村居民所占参与训练数据比例分别为 6.765％和 50.529％，要明显高于节点 2 中的 0 和 23.256％。这说明收入因素是影响农村居民总体满意度的最主要因素。在诸多叶节点中，我们选取具有代表性的两个节点进行比较。节点 15 反映了对自身收入、所居住村社会风气、居住环境非常满意的人群，总体满意度最高，此类人群在总体满意程度问项中选择“非常满意”的农村居民所占参与训练数据比例达到了 92.057％；而节点 6 反映了对自身收入、所居住村社会风气非常不满意的人群，总体满意度最低，此类人群在总体满意程度问项中选择“不满意”的农村居民所占参与训练数据比例达到了 90％。

7. 模型评估

模型建立后，需要对模型进行验证和评价，以得到更合理、更完备的决策信息。通过模型评估可以确定前面的各步骤是否已经成功地完成，所以其是数据挖掘过程中的关键一步。下面主要采用建立模型收益图和计算验证集数据结果进行对比来进行模型评估。

（1）收益图

收益图的对角线表示的是如果不使用模型，整个样本的期望响应，而收益曲线越陡，增益就越高。如果不使用模型，调查人数与所获得的收益成同比例增长，即在调查对象中，反映为“比较满意”的人群占总人群的比例与所调查的人群比例相等。图 9-6 即为模型收益图，图中曲线表示，随着样本量的增加，基于增益能多大程度地提高响应率。以深度细分模型中总体满意度为“比较满意”为例，由模型收益图可以看出，本研究设定的模型对于修正总体具有一定的提升意义。

（2）通过验证集检验模型准确性

我们在固定随机数种子为 177 216 的前提下，在 Sample 节点中，

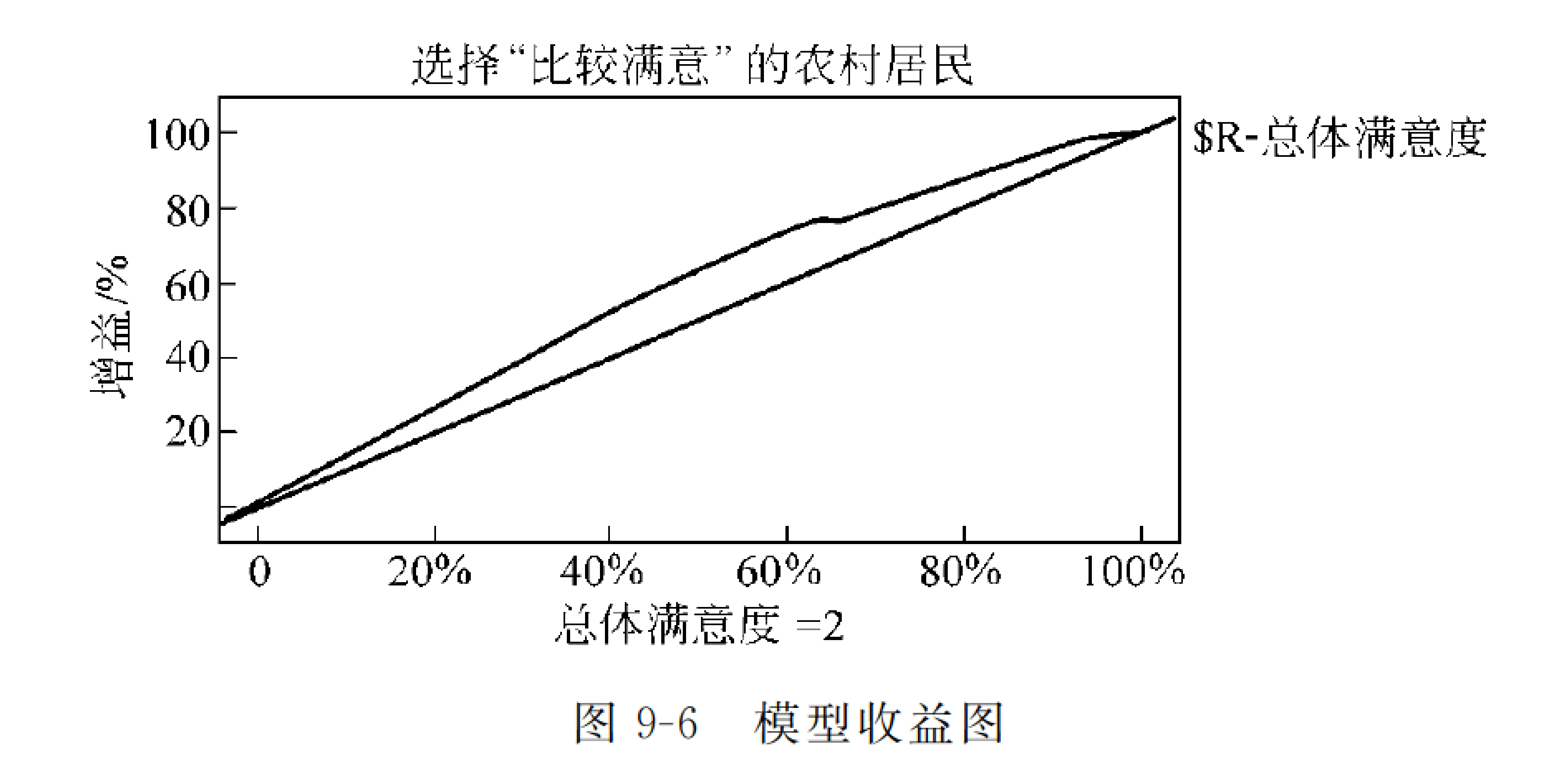


图 9-6　模型收益图

选取“丢弃样本”即抽取由剩下的 15%的数据组成的验证集进行模型验证,并且通过矩阵节点观察验证结果。广度细分模型准确率和深度细分模型准确率见表 9-5。

表 9-5　模型准确性

类别	广度细分验证结果		深度细分模型结果	
	个案数	比例/%	个案数	比例/%
正确	63	58.33	83	76.67
错误	45	41.67	25	23.33

从分析结果可以得出,验证集中共有 108 条记录,当用验证集对广度模型进行检验时,发现 108 条记录中的 63 条被正确预测,占总数的 58.33%;而另外的 45 条记录被错误预测,占总数的 41.67%。当用验证集对深度模型进行检验时,发现 108 条记录中的 83 条被正确预测,占总数的 76.67%;而另外的 25 条记录被错误预测,占总数的 23.33%。两个模型预测的正确率相对较高,但是没有达到 80%以上,收集到的有效数据记录只有 905 条,样本量较低是导致此问题的主要原因。假设在能够获得充足样本量的前提下,这两个检验的准确度将会得到进一步改善。广度细分模型和深度细分模型正确预测个数矩阵为 $\boldsymbol{\omega}_1$ 和 $\boldsymbol{\omega}_2$,其中对角线上元素为预测准确的个数:

$$\boldsymbol{\omega}_1=\begin{pmatrix}0&0&0&0\\1&60&21&23\\0&6&3&0\\0&0&0&0\end{pmatrix},\quad \boldsymbol{\omega}_2=\begin{pmatrix}7&1&0&0\\1&52&9&0\\0&6&20&0\\0&2&6&4\end{pmatrix}$$

由模型预测矩阵可以看出，模型 1 正确预测的 63 个个案中，有 60 个被测者选择了“比较满意”，有 3 个被测者选择了“不太满意”；模型 2 正确预测的 83 个个案中，有 7 个被测者选择了“非常满意”，52 个被测者选择了“比较满意”，20 个被测者选择了“不太满意”，4 个被测者选择了“不满意”。总体来说，模型预测准确性尚可。

9.3　新农村建设满意度细分模型的结果分析

由模型的结果可以得出，村庄类型因素对于农村居民总体满意度的影响力度最大；性别因素对于农村居民总体满意度的影响力度最小，几乎无影响。所在村庄类型为新农村建设和新民居建设示范村的农村居民，普遍满意程度较高，总体满意程度达到了 66.078%，新农村达标村、移民搬迁村和其他类型村庄的农村居民总体满意度很低，总体满意程度仅有 46.234%。这说明河北省新农村建设和新民居建设取得了一定的成效，村庄结构的改造为农村居民带来了一定的福利。另外，结果发现，文化程度趋中者（主要群体为小学、初中文化者）总体满意程度略低于不识字或少识字农村居民群体和高中及高中以上文化群体。在调研时对河北省农村居民的访谈过程中我们发现，文化程度较低者比较安于现在的生活状态，而文化程度较高者多为在县城上班，工作地点不在居住村庄，故对于当前生活满意度较高。而文化程度趋中者，具有一定的能力，但是现阶段由于农村集体经济发展力度不够，在村庄中无法为其提供满意的工作或职务。这一方面反映了当前河北省新农村建设和新民居建设应当加大力度；另一方面也体现了新民居建设的重要性——大力实行新民居建设，可以促进村庄之间的聚合，集中力量办大事，产业形成规模发展，从而使农村居民中具有一定文化程度者有用武

之地。另外，收入依然是影响其整体满意度的关键。另外，社会风气因素、居住环境因素对于农村居民总体满意度也有着不小的影响。因此，政府在制定新农村建设、新民居建设政策时，要协调发展，统筹兼顾，以民居建设模式来带动农村产业的发展，从而带动农村社会风气和居住环境的改善。

第 10 章　基于关联规则的新农村建设满意度影响因素分析

10.1　关联规则简介及模型选择

关联规则是数据挖掘中的一种主要的挖掘技术，也是非常活跃的研究领域，由 Agrawal、Imielinski 和 Swami 于 1993 年首先提出。关联规则挖掘擅长于发现大量数据中项集之间有趣的关联或相关联系。其最早产生于发现购物篮数据中隐含的模式。随后吸引了许多研究者的兴趣，其理论不断深入和完善，应用领域也不断扩大。除了购物篮数据外，还可以应用于其他领域。不但能够挖掘交易数据库，而且对数值型数据和分类数据都能够进行很好的分析和处理。

关联规则是形如 $A \Rightarrow B$ 的蕴含表达式，A 为规则的前件，B 为规则后件。其中 A 和 B 是不相交的项集，即 $A \cap B = \varnothing$。关联规则的强度可以用它的支持度和置信度度量。支持度确定规则可以用于给定数据集的频繁程度，而置信度确定 B 在包含 A 事务中出现的频繁程度。支持度与置信度的度量形式定义如下：

$$\text{Support}(A \Rightarrow B) = \frac{\sigma(A \cup B)}{N} \tag{10-1}$$

$$\text{Confidence}(A \Rightarrow B) = \frac{\sigma(A \cup B)}{\sigma(A)} \tag{10-2}$$

大多数关联规则算法采用两步挖掘策略，首先寻找所有满足最小支持度阈值的频繁项集，接着从第一步所发现的频繁项集中提取所有具有高置信度的规则，通过这两步挖掘出来的规则称为强规则。

本章采用支持度-置信度框架来规定关联规则。另外，一般情况下，有用的关联规则的作用度应该大于 1。因此，在规则的选取中，本

章把作用度作为关联规则评价的另一个标准。

对于关联规则挖掘目前常用的有 Apriori 算法、GRI 算法和 Carma 算法。其中 Apriori 算法通常挖掘的是布尔关联规则，它考虑的是项的在与不在。对于标称或量化等属性变量类型数据进行关联规则挖掘时，部分文献将非布尔类型变量转换成布尔变量类型以适合 Apriori 算法。而 GRI（广义规则归纳）算法是从数据中提取一组规则，找出信息容量最高的规则。与 Apriori 算法相比，它既能处理标称属性输入变量又能处理量化属性输入变量，还能同时处理多个输出变量。

本问卷分析过程中用到的关联规则主要分为两个方面：第一个方面是农村居民对于新农村建设和新民居建设影响因素分析，由于此部分后项变量为多个输出目标，信息处理量较大，因此采用 GRI（广义规则归纳）关联规则算法；第二个方面是农村居民对于目前认为所存在的问题的解决途径分析，前项变量和后项变量均为特征变量（选项定义为 0—未选则，1—选择），故应用 Apriori 关联规则算法更为合适。

10.2　基于 GRI 算法的关联规则建模

1. 数据准备及模型调整

（1）定义数据源

设定关联前项变量属性共有 11 个，关联后项基本变量 3 个，经过数据预处理部分，得到处理之后的数据集“关联分析数据 1. xls”，并将此数据定义为数据源。

（2）关联规则前项设置

由上述关联规则相关原理发现，当某些属性变量某个标志值样本个数不够频繁时，可能会对关联规则结果造成影响，丧失一些可以挖掘的知识，造成某些频繁模式指数的增长，增大计算开销。因此，本研究在数据与处理部分将原始数据分割出来的用于关联规则的数据进行了以下处理（具体反映在节点中，不再赘述）。根据决策树分析结果和样本描述统计量，将样本数量较少的标志或者样本结构类似的标志进行合并，主要变化为：将“年龄”变量中的 30 岁以下（不含 30 岁）和 30～

45 岁合并成低龄组，45～60 岁（不含 60 岁）和 60 岁以上合并成高龄组；将“文化程度”变量中不识字或少识字者和小学文化者合并为低文化程度组，初中为中等文化组，高中和大专以上（含大专）为高文化程度组；将“职业”变量中的村干部、乡镇干部和其他分为一组；将“政治面貌”中群众分为一组，剩余者成为另一组；将“所在村类型”变量中示范村类型为一组，剩余的移民村、达标村和其他村合并为另一组；将“居住类型”变量中自建小楼和居民区类型合并为一组，作为新民居建设模式组，另一组由其他居住模式合并，作为传统民居模式组。变量类型设置根据变量自身特征完成。具体变量类型设置和对应问卷代号等信息详见表 10-1。

表 10-1　关联规则前项变量表

变量名	变量对应问卷内容	对应问卷代号—选项	类型
A_1	性别	1—男；2—女	标称
A_2	年龄	1—低龄组；2—高龄组	标称
A_3	文化程度	1—低文化程度；2—中等文化程度；3—高文化程度	数值
A_4	职业	1—农民群众；2—村干部及其他	标称
A_5	政治面貌	1—党员和民主党派；2—群众	标称
A_6	所在村类型	1—示范村；2—非示范村	标称
A_7	居住类型	1—传统民居模式；2—新民居建设模式	标称
C_5	收入	1—非常满意；2—比较满意；3—不太满意；4—不满意	数值
C_{11}	居住环境	1—非常满意；2—比较满意；3—不太满意；4—不满意	数值
C_{15}	社会风气	1—非常满意；2—比较满意；3—不太满意；4—不满意	数值
C_{17}	领导干部	1—非常满意；2—比较满意；3—不太满意；4—不满意	数值

（3）关联规则后项设置

本研究的此部分主要目的是发现河北省新农村建设和新民居建设对于农民感知的影响因素和影响程度，因此，选取问卷中与河北省农村居民对于政策理解和政策支持程度相关的三个变量作为关联规则后项变量。其中，在数据预处理部分将“对于新民居政策的态度”变量中的问项进行合并，同样是为了避免不频繁问项对于结果的影响。其余两个变量问项不作此类处理。类型设置根据变量自身特征完成（同上）。

具体变量类型设置和对应问卷代号等信息详见表 10-2。

表 10-2 关联规则后项变量表

变量名	变量对应问卷内容	对应问卷代号—选项	类型
B_1	开展新农村建设与自己	1—密切相关;2—不相关;3—不清楚	标称
B_2	新农村建设政策了解	1—知道;2—知道一点;3—不知道	标称
B_6	如何看待新民居政策	1—肯定态度;2—中立态度;3—否定态度	标称

(4) 阈值设定,反复挖掘

经过预处理后的数据由 Clementine 12.0 数据挖掘软件中的 GRI 模型来实现挖掘过程。在关联规则挖掘中需要设置相关阈值。其中,最小支持度和置信度阈值的大小直接影响着关联规则挖掘出来的规则的数量。如果对最小支持度和置信度阈值不加限制,产生的规则数量将会十分惊人,结果将毫无意义。相关文献中并没有明确的依据确定如何设置模型的最小支持度阈值和置信度阈值,本研究在实际操作中,采用设置不同最小支持度阈值的方式进行反复挖掘,即进行若干次探索性的挖掘,由结果可知,如果最小支持度阈值设得太低,有可能产生大量冗余规则;如果最小支持度阈值过高,则很可能有价值的关联规则信息,最小支持度和置信度阈值的调整明显影响着关联规则的数量。例如,如果将置信度设为 30%,当最小支持度阈值为 1%时有 3921 条规则,置信度设为 50%,当最小支持度阈值为 3%时,后件为$\{B_2=1\}$的所有模式都将损失掉,部分结果见表 10-3。

表 10-3 不同支持度和置信度阈值情况下的关联规则数量表(B_2)

	置信度	30%				50%			
	后件	总数	1	2	3	总数	1	2	3
支持度	1%	3921	1685	1012	1224	305	50	61	194
	3%	1125	276	580	269	14	0	8	6
	10%	241	4	224	13	0	0	0	0
	15%	97	0	96	1	0	0	0	0

2. 关联规则结果

经过多次比较，本研究最终选择分析在最小支持度阈值为 10%、最小置信度阈值为 30%的情况下进行关联规则挖掘。下面对各个后项变量进行分析。从诸多作用度大于 1 的关联规则中按照最小支持度进行排序，选出排名靠前且有代表性[①]的规则，由于规则众多，在确保支持度大于 10%的条件下，选取部分置信度大于 40%的规则进行论述，选出的规则详见表 10-4、表 10-5、表 10-6。

表 10-4　后项变量为 B_1 的关联规则(部分)

规则	前　　项	后项	支持度/%	置信度/%
1	$A_3>2.5$	1	36.28	43.06
2	$A_6=1\wedge A_7=2$	1	25.17	46.21
3	$A_6=1\wedge C_{11}<2.5\wedge C_{15}<2.5\wedge A_7=2$	1	16.67	51.04
4	$C_5<3.5\wedge A_4=1\wedge A_5=2\wedge C_{15}<2.5\wedge A_7=1$	2	18.06	63.46
5	$A_1=2\wedge A_7=1\wedge C_{17}<2.5$	2	15.8	64.84
6	$A_2=2\wedge A_6=2\wedge C_{17}<2.5$	2	11.11	65.62

表 10-5　后项变量为 B_2 的关联规则(部分)

规则	前　　项	后项	支持度/%	置信度/%
1	$A_6=2\wedge C_{15}<2.5$	2	30.03	56.07
2	$A_6=2\wedge C_{15}<2.5\wedge C_{17}<2.5$	2	23.09	60.15
3	$C_{15}>2.5\wedge C_{17}>2.5\wedge C_5>2.5\wedge A_5=2$	3	12.33	70.42
4	$C_{15}>2.5\wedge A_3<2.5\wedge A_4=1\wedge A_5=2$	3	12.5	72.22

(1) 后项变量为 B_1 的关联规则

如表 10-4 所示，规则 1 说明文化程度相对较高的人群中，认为新农村建设与自身密切相关，其概率为 0.4306；规则 2 说明，在新民居建

① 由于本问卷调研的目的在于发现新农村(新民居)建设的成效和不足之处，因此，对于规则前项中含有居住类型、是否为新民居建设示范村最为关注，另外，对文化程度、各个方面满意度进行适当关注。

表 10-6　后项件变量为 B_6 的关联规则

规则	前　　项	后项	支持度 /%	置信度 /%
1	$A_6=1 \wedge C_5<2.5$	1	32.99	74.74
2	$A_4=1 \wedge C_{11}<2.5 \wedge C_{17}<2.5 \wedge C_5<2.5 \wedge A_6=1$	1	20.31	79.49
3	$C_5>2.5 \wedge A_5=2 \wedge A_7=2$	3	10.07	41.38

设示范村和新农村建设力度较大，住房模式进行改造的村庄中的居民对于新农村建设的期望很大，认为新农村建设与自身密切相关的概率为 0.4621。这也说明了新农村建设和新民居建设取得了成效，使农民认识到了新农村建设与自身的密切联系。规则 3 则进一步说明了在规则 2 中所定义的居民中，对于社会风气和居住环境比较满意的人群中认为新农村建设与自身密切相关的概率为 0.5104；规则 4 说明了在未进行新民居建设的村庄中，对于收入和社会风气比较满意的居民中，认为新农村建设与自己不相关的概率为 0.6346；规则 5 和规则 6 则反映了在新农村建设推行缓慢、未开展新民居建设的村庄居民中，对于村领导干部比较满意，此类人群中女性受访者和年龄较大者认为新农村建设和新民居建设与自己没有关系的概率分别为 0.6484 和 0.6562。

(2) 后项变量为 B_2 的关联规则

如表 10-5 所示，规则 1 说明非新民居建设示范村的人群中且对社会风气比较满意者，对于新农村建设相关政策不太了解，其概率为 0.5607；规则 2 说明，在规则 1 所确定的人群中，领导干部比较满意者对于新农村建设相关政策不太了解的概率为 0.6015；规则 3 则说明了对于收入、领导干部、社会风气未达到满意的人群中不了解新农村建设政策的概率为 0.7042；规则 4 说明在普通群众中文化程度较低且对于社会风气未达到满意的人群不了解新农村建设政策的概率为 0.7222。

(3) 后项变量为 B_6 的关联规则

如表 10-6 所示，规则 1 说明新民居建设示范村的人群中对于自身收入满意者且对新民居建设持肯定态度的概率为 0.7474；规则 2 说明，在规则 1 所确定的人群中，对于居住环境、村领导干部比较满意者对新民居建设持肯定态度的概率进一步提高，达到了 0.7949；规则 3

则说明了新民居建设示范村的农民群众，收入未达到满意者对于新民居建设持否定态度的概率为 0.4138。

10.3　基于 Apriori 算法的关联规则建模

1. 数据准备及阈值设定

本次调研问卷的问项设置另外涉及两大部分，一是“建设新农村的主要困难”，进一步分为缺乏资金原因，农民致富难，农民就业难，发展现代农业难，发展农村二、三产业难，壮大村级集体经济难，农民素质提高难，构建“工业反哺农业，城市扶持农村”长效机制难，缩小城乡差距难九个部分；二是“对新农村建设的企盼”，进一步分为开展新农村建设应从何处着手、对新农村建设有何期盼、通过哪些渠道实现早日致富、应该如何发展现代农业四个方面。每个方面构成一个问题，每个问题中有若干选项，在问卷录入时，我们设置每个选项作为一个变量，则第一部分对应的变量数为 43 个，第二部分对应的变量数为 33 个。以“通过哪些渠道实现早日致富”为例，具体变量设置见表 10-7。每一个变量均只取到两个值，即如果选取此选项，记为 1，未选取记为 0。这样，这些变量便满足了 Apriori 关联规则挖掘的要求。在 Clementine 12.0 节点中，设置前 43 个变量为前项变量，后 33 个变量为后项变量，进行关联规则挖掘。由于题项众多，为了避免过于繁多的规则被发现，经过多次比较，我们最终选择分析在最小支持度阈值为 20%、最小置信度阈值为 50%的情况下进行关联规则挖掘。

表 10-7　题项“通过哪些渠道实现早日致富”细分变量设置

问项内容	变量设置	录入方式
通过有偿流转获得更多的土地或养殖更多的牲畜，实现规模经营	T_1	1—选取；0—未选取
获得更多的技术或信息支持	T_2	1—选取；0—未选取
获得稳定的农字号龙头企业或专业协会的带动保障	T_3	1—选取；0—未选取

续表

问项内容	变量设置	录入方式
获得稳定的就业渠道	T_4	1—选取；0—未选取
通过参加各类培训提升自己的综合素质	T_5	1—选取；0—未选取
获得平等的国民待遇	T_6	1—选取；0—未选取

2. 关联规则结果

将所有作用度大于1的规则中，对支持度大于20%且置信度不低于50%的所有规则进行了汇总。结果发现，虽然我们设定的后项变量有33个，但是产生规则的变量集中于7个变量，具体情况见表10-8。

表10-8　规则汇总表

后项内容	相关规则数	规则最小置信度/%	规则最小支持度/%
获得更多的技术或信息支持	269	60.086	20.139
获得稳定的就业渠道	128	50.00	20.66
通过参加各类培训提升自己的综合素质	46	50.00	20.486
无论种植还是养殖要实现规模经营	95	50.00	20.139
必须依靠科技进步，加快农业科技的推广应用步伐	281	61.24	20.312
因地制宜发展特色农业	19	50.00	20.139
有科技带头人的示范作用	149	50.00	20.312

计算得到的规则众多，由于篇幅问题不再单独一一列举。观察规则前项可知，43个前项变量在不同的规则中，对于后项“获得更多的技术或信息支持”和“必须依靠科技进步，加快农业科技的推广应用步伐”均有体现，而规则数最少的后项变量“因地制宜发展特色农业”，其前项变量数也涵盖了43个前项变量中的16个。由此可见，在这一部分的关联规则挖掘中，我们得到的规则相当多，但是又能从中发现的规则信息却不多。但是我们可以总结得出的是，当前新农村建设中，需要解决

一系列问题，而关键在于农村科技的进步和农民素质的提高。

10.4　基于关联规则的新农村建设满意度影响因素分析

关联分析结果显示，村中是否有整体规划会影响人们是否希望参与新民居建设。一方面，如果村里没有统一规划，而村民的居住环境已经满足了人们的愿望，可能暂时并不想尝试新事物，对新农村建设和新民居建设关注程度较小；另一方面，正是由于村里有统一规划，农民看到了统一规划的好处，享受到了统一规划给自己生活带来的方便，可能会希望得到更加完善的服务和设施，参与新民居建设的积极性更高。另外，村中没有整体规划的被调查者中，对各个方面不甚满意的人群，期望统一规划村庄布局。从人群自身特点来看，被调查者的文化程度越高，对新农村建设政策理解程度越高，越希望参与新民居建设。研究结果表明，农民的文化程度、所在村庄是否进行了新民居改造，是影响农民意愿和政策理解的重要因素。这就要求政府增加对农民的教育投资，进一步提高农民的文化程度，改进农村人力资本状况，从而提高农民响应政府政策的积极性。结果还显示，在农村居民普遍关注的问题当中，最重要的方面在于科学技术的掌握和获得稳定的就业渠道。因此，政府在以后的农村建设工作中要加大科技下乡力度，带动农村产业发展，为农民提供更广阔的发展空间，拓宽就业渠道，从而提高全体农民的收入水平。

第 3 篇

实现农村现代化
提升农民生活质量

第 11 章　农村现代化的界定及指标体系建立

11.1　农村现代化的界定

农村现代化将从农村整个社会体系进行研究，便于从整体把握农村的现代化的发展和进程。现代化理论认为，社会的发展是改变之前的传统社会，这需要一个长久的过程。为了达到此目标，就应先实现全面农村现代化，这样整个社会才能综合发展。同时，农村的现代化不仅仅是农村社会的稳定，同时也要重视农民的生活水平，只有将物质、精神两方面都纳入研究范围，才能系统地考察农村现代化。从农村的整体发展来看，农村现代化是农民利用各种条件，从政治到经济、从文化到思想全面发展的过程。农村现代化要求的不仅仅是能够满足农民的物质生活，经济市场繁荣，也需要社会安定、生活富裕；此外，现代化的农村，是农村全面实现农业机械化、经营产业化、结构非农化的农村，也是乡村城镇化、生活现代化、农民知识化、社会文明化的农村。

故此，我们将农村现代化定义为：经济繁荣、科技进步、文教发达、生活富裕、文明和谐、社会安定、环境优美，农村全面实现农业机械化、经营产业化、结构非农化、乡村城镇化、生活现代化、管理科学化、保障社会化、农民知识化、社会文明化。

11.2　指标体系的建立

根据上一节的定义以及研究的需要，本节将农村现代化的要求归类为农业生产方面、农村生活方面和农民素质三个大方面的现代化，如

图 11-1 所示。

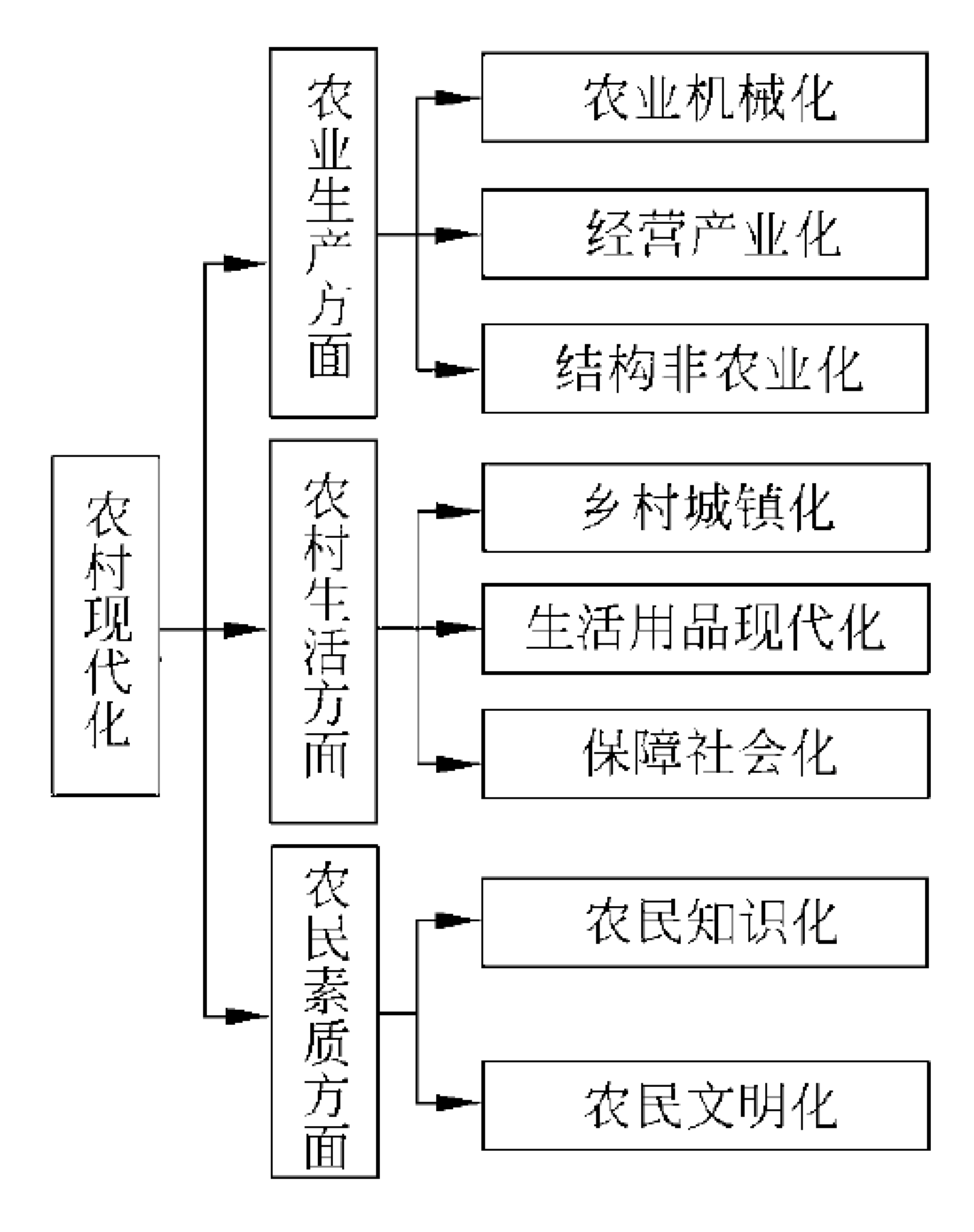


图 11-1　农村现代化指标体系

农业生产方面中，现代化包含农业机械化、经营产业化、结构非农业化，这些都是农村现代化的基础。如果没有生产现代化，那么农村现代化也就失去了扎实的物质基础，农村现代化也将是空谈。因此，在农村现代化的过程中，不仅要把传统生产逐步改造为商品化，更要实现产业化、技术化、社会化、生态化、国际化。没有现代化的农业，就不会有现代化的农村。

农村生活方面现代化包含乡村城镇化、生活用品现代化、保障社会化；农村的现代化最终要反映在农民生活上，农村现代化的核心即为推进农村经济的发展，推进农村现代化的目的也为切实提高农民生活水平，让农民享受更多的经济发展成果，即提高农村生活方面实现现代化。

农民素质方面现代化主要是农民知识化和文明化。农村现代化的建设主要取决于农村建设的主体——农民，因此，农民素质现代化的地位不容小觑。没有现代化的农民，就没有现代化的农业和现代化的农村。因此，在农村现代化的过程中，传统意义上的个体农民应当逐步转变为用现代科学技术武装起来的以市场为导向的日趋文明化的现代职业群体。

第 12 章　基于决策树的农村现代化对农民生活满意度影响分析

12.1　变量的选择

问卷得到的数据含有 167 个变量，其中有许多变量比如对生活状态的满意度与收入满意度有较强的相关性，有的甚至有重复性，而我们是要找出对收入满意度有隐含的决定性质的变量，故不再选用。

还有一些变量是由农民收入所决定的，比如家里电器使用、交通工具使用的情况等，根据决策树的算法，此类变量必将大大影响决策树的质量，因此也予以剔除。

此外还有一些与收入无关的干扰变量也过滤掉，如对农村教学条件是否满意、是否有大学生村官等。

因而最终选择的变量有 B_3 养殖业情况、B_5 收入来源、B_{16} 国家粮食补贴政策的情况、B_{19} 对畜牧业补贴政策是否满意、C_1 所在村种植业情况、C_2 所在村耕地机收情况、C_3 农资供应情况、C_4 自然条件、C_6“六通六化”建设满意情况、C_{14} 对医疗养老社保满意度、C_{18} 对新农合满意度、C_{19} 对农村养老保险制度满意度、C_{20} 对最低保障制度满意度、C_{21} 对一池三改满意度、C_{22} 对移民搬迁工程满意度、$I_1 \sim I_4$ 缺乏资金的原因、$J_1 \sim J_8$ 农民致富的主要困难、$K_1 \sim K_3$ 农民就业主要困难、$O_1 \sim O_3$ 农民素质提高难的原因、$R_1 \sim R_{15}$ 新农村建设从何着手、$S_1 \sim S_7$ 有何期盼、$T_1 \sim T_6$ 通过哪些渠道早日致富。

12.2 农村现代化对农民生活满意度影响的决策树模型

将新农村建设示范村的调查问卷处理后数据导入 Clementine，将农民收入满意的变量 C_5 设为输出变量，除编号外都为输入变量，接分层节点。由于新农村建设示范村的有效问卷共有 576 份，为更有效地进行训练，故将训练样本集同测试集的比例设为 80 ∶ 20，然后链接 C5.0、C&R 等模型进行建模。

12.2.1 基于 C5.0 算法对所有可能影响因素建模

C5.0 以信息增益率为标准确定最佳分组变量和分割点。对于候选属性集中的每个属性，C5.0 算法首先计算该属性上每种可能划分的信息增益率，然后选择信息增益率最大的划分作为该属性的最佳划分，最后比较所有候选属性最佳划分的信息增益率，选择具有最大划分信息增益率的属性作为最终测试属性。C5.0 同时使用前剪枝和后剪枝两种修剪方法来构建决策树。

C5.0 算法中 Boosting 技术大幅提高了算法的分类准确度，使 C5.0 算法在处理大规模连续变量数据集上更具有优势。Boosting 是 C5.0 算法的一种集成技术，其最大的特点是生成一组分类器。对于所有的训练样本集，每个样本集都被赋予一个权重，如果某个样本集被准确分类，那么构造下一个训练集时，它被选中的概率就会降低；如果某个样本集分类错误，那么它的权重就会提高。通过这种方法，Boosting 技术通过递归，使用更新过的权重的训练样本集来训练下一个分类器，提高了算法的分类准确度。

应用 5.0 节点进行编辑。由于 Boosting 技术可以提高模型预测的稳健性，我们勾选了 Boosting 技术，试验次数为默认的 10。Group Symbolics 选项会在分支前看分组变量是否可以合并，可以使得到的树比较精简。我们选择剪枝策略设为 Expert，首先都按默认的运行，即设定精度为 75%，每个节点上所允许的最小样本量为 2。

设定损失矩阵。实际上的满意度模型预测时将满意判别为不满意时，信息损失较小，但若把不满意的预测成为满意的则将影响分析结果，故设定的损失矩阵表如表 12-1 所示。之后对模型反复试验剪枝策略，运行程序，得出表 12-2。

表 12-1　损失矩阵

类　别		预　测	
		满意	不满意
实际	满意	0	4
	不满意	8	0

表 12-2　不同参数下 CART 模型拟合度

差异性测度指标	修建树	使用标准误规则乘数	训练集正确率/%	测试集正确率/%	深度
无	无	无	85.08	76.29	7
Gini/Twoing/Ordered	有	1.0	82.72	74.23	7
Gini/Twoing/Ordered	无	无	82.46	72.16	5
Gini/Twoing/Ordered	有	1.0	79.58	70.62	5

从图 12-1 以及运行的结果可以看出，模型的拟合度和信息增益率都较高。从模型得出变量的重要性，前 10 名依次为："六通六化"建设的满意度、农资供应满意度、农民收入满意度、农产品价格低且不稳定、养殖业情况、通过获得更多的技术或信息支持早日致富、缺少资金人才、养殖方式、兴水沃土工程建设。将"工业反哺农业，城市支持农村"机制落在实处。

因此我们推断，为增加农民生活满意度，村中"六通六化"等基本工程需要扎实落实，农资供应要充足，新农村建设还要多从各种角度促进农业增收，村中多帮助农民提高养殖业和种植业，可以先选出带头人做好增收的榜样，然后再传授经验帮助他人。此外，还需加大医疗养老社保的受惠群众。农民希望获得龙头企业或专业协会的带动保障，从而增加收入。

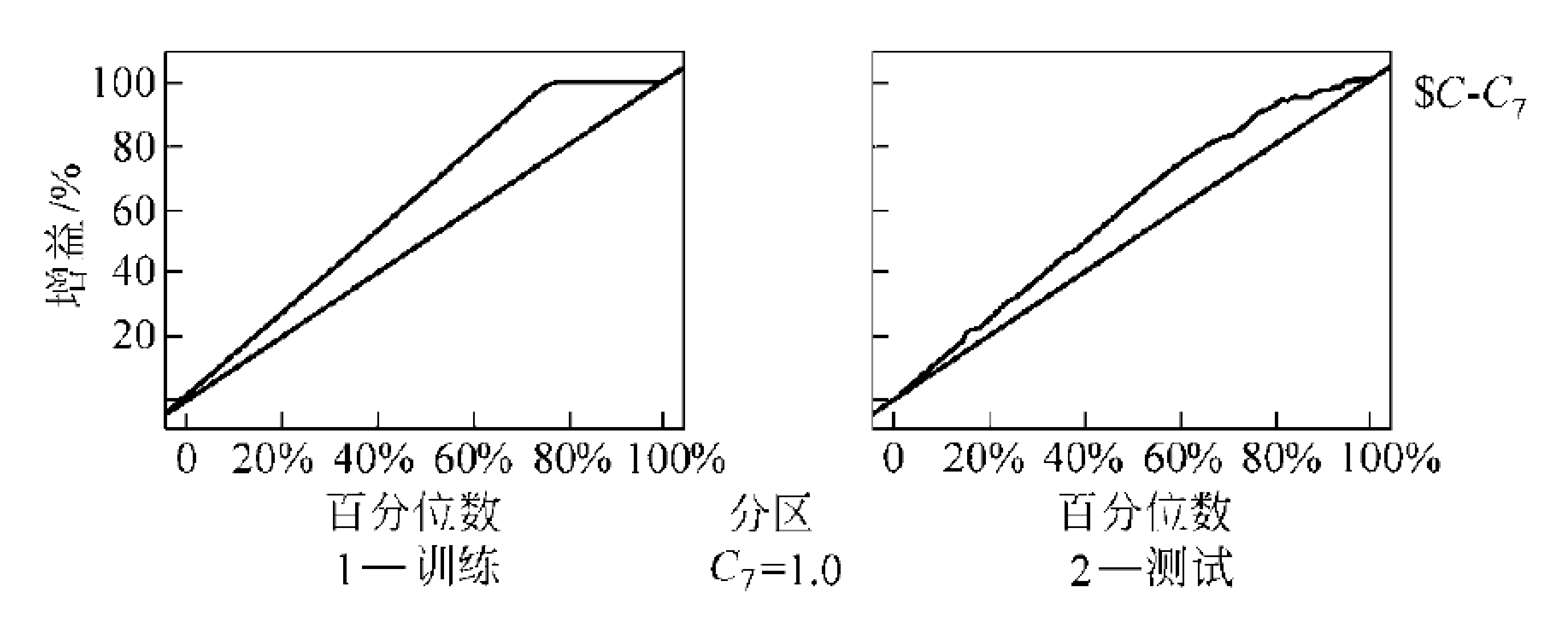


图 12-1 C5.0 模型对农民生活满意度的收益图

12.2.2 基于 CART 算法的生活满意度的农业生产因素分析

在上一部分，虽然 C5.0 模型拟合较好，但变量太多且决策树过于茂盛。为进一步研究村落生产因素对农民生活满意度的影响，先仅选取生产变量，选用的变量有农民对种植业情况评价、机耕机收的情况评价、农资供应评价、养殖业情况评价、养殖方式、对工业反哺农业政策实施情况的评价、对畜牧业补贴政策的满意度、缺少资金的原因判断，农民致富困难的主要原因和致富的方法等，进一步用 CART 算法拟合模型。

CART 算法是一种基于二叉树的分类预测方法。CART 算法利用二分递归分割的技术，每次将当前样本集划分为两个子样本集，使得生成决策树的每个非叶节点都存在两个分支。可对高度倾斜和多态的数值型数据，以及顺序或无序的类属性数据进行处理。

对于候选属性集中的每个属性，CART 算法通过计算该属性每种可能划分的 Gini 指标，找出 Gini 指标最小的划分作为该属性的最优划分，最后比较所有候选属性中的最优划分 Gini 指标，具有最小划分 Gini 指标的属性成为最终的测试属性。该方法可使用递归分类方法将训练样本集分裂为具有相似输出字段值的段别。通过检查输入样本集找到最佳分裂点，并根据由该分裂点所导致的“纯度”降低情况进行测试，从而选出该分裂点的最佳划分。

CART 算法考虑到每个节点都可能成为叶子节点，对每个节点都分配类别，分配类别的方法可以用当前节点出现最多的类别，也可以参

照当前节点的分类错误或者其他更加复杂的方法。针对样本数比较少而不能分出独立的测试样本集的情况，CART 算法采用了一种称为交叉确定(cross validation)的后剪枝方法，该方法的优点在于使用尽可能多的训练记录，能够有效地覆盖整个数据集，从而解决小样本集决策树没有独立测试样本集而造成的过度拟合问题。

我们选择后剪枝方法，设定损失矩阵，设定不同参数建模，得到不同参数下的 CART 模型拟合度如表 12-2 所示。

从表 12-2 中我们可以看出，差异性测度指标、修剪树并使用标准误规则乘数这几个参数所带来的模型拟合度差异并不大。因此我们选取训练集正确率次高，但测试集正确率最高，深度为 7 的简单模式进行建模。得到村落变量重要性排名如下。

第一层，村民对畜牧业补贴政策实行情况的满意度。在生产因素中，畜牧业补贴政策实行情况的满意度对农民生活满意度的影响最大，为 0.17。对畜牧业补贴非常满意和满意的农民有 80%以上对生活状况满意，对畜牧业补贴不太满意和不满意的则有一部分对生活满意，一部分对生活不太满意。这说明，一般来说，畜牧业补贴对农民生活帮扶力度大，能有效改善农民生活。

第二层，村民对村中种植业情况评价。在生产因素中，村中种植业情况对农民生活满意度的重要性为 0.12。村民所在村种植情况好的对生活满意的居多，而一般的则对生活的满意度不定。这是由于河北省农村生活现阶段主要还是依靠小规模农业为主，大部分留守在村中的还是依靠种植业生活，故村中种植业较好的，农民对生活也较满意；而种植业不太景气的村落，部分农民依靠在工厂务工或个体经商，部分依旧务农，故对生活的满意度参差不齐。

第三层，农资供应情况、所在村耕地机收情况、缺乏人才、兴水沃土工程建设等变量的重要性都较小，其对农民生活满意度的影响较弱。

12.2.3　基于 CART 算法的生活满意度的农村生活因素分析

为研究影响农民生活满意度的生活因素，重新选择变量，参数定位上一节拟合度最高的模式，即使用差异性测度，修剪树，并使用 1.0 的标准误规则乘数，其余默认，进行建模，得到决策树如图 12-2 所示。

从图 12-2 可以看出在只有生活状况因素下,影响村民对生活满意度的因素排在最前面的便是村民对村里“六通六化”的满意度,超过了 0.2。这说明村里的通水、通电、通路、通电话、通班车等基本情况影响着村民的收入水平,即对这些基本情况较满意的说明所在村的发展水平达到村民心中的基本要求,而对情况不满意的说明所在村的发展水平相对于周边村要差一些,收入水平也低一些,导致村民对收入的满意度不高。

其次,是最低保障制度满意度对生活状况满意度,变量重要性为 0.15。这说明最低保障若执行有力,可以给村民减少相当程度的生活负担,能大大提高农民生活的满意度。对低保不满意的群体,生活负担依然沉重,渴望获得低保,或获得较高保障水平,对生活满意度不高。

然后,收入来源对生活满意度重要性位于第三位。结合模型可以看出村民收入来源为农业、养殖业和务工的对收入满意的比例不如乡镇企业工作、做生意或其他。

再次,为农民对新型合作医疗的满意度、养老社保的满意度。医疗养老等是农民的重大开支之二,新型农村合作医疗和养老保险的逐步实施,落户农村,将给农民减轻极大的负担。

12.2.4 基于 C5.0 算法的生活满意度的农民素质因素分析

为研究影响农民生活满意度的思维方式因素,重新选择变量,参数定为 12.2.2 节拟合度最高的模式,即使用差异性测度,修剪树,并使用 1.0 的标准误规则乘数,其余默认,进行建模,模型增益图见图 12-3。测试集正确率仅为 63%。再反复试验得到用算法默认简单模式建模得到的正确率较高,为 76%,但将对生活状况不满意的判别成满意的失误较高,故现采用 C5.0 算法进行建模,得到模型正确率为 77%,因此最终选用 C5.0 算法对个人因素进行分析。

分析结果显示,变量重要性排名为:对教育环境的满意度、希望子女受教育条件改善之处、农民素质提高难因之培训太少、希望大学生村官带来的帮助、对教学条件的满意度、最关心的教育问题、对提高农民素质培训的满意度、农民素质提高难因之认识不到知识重要、农民素质

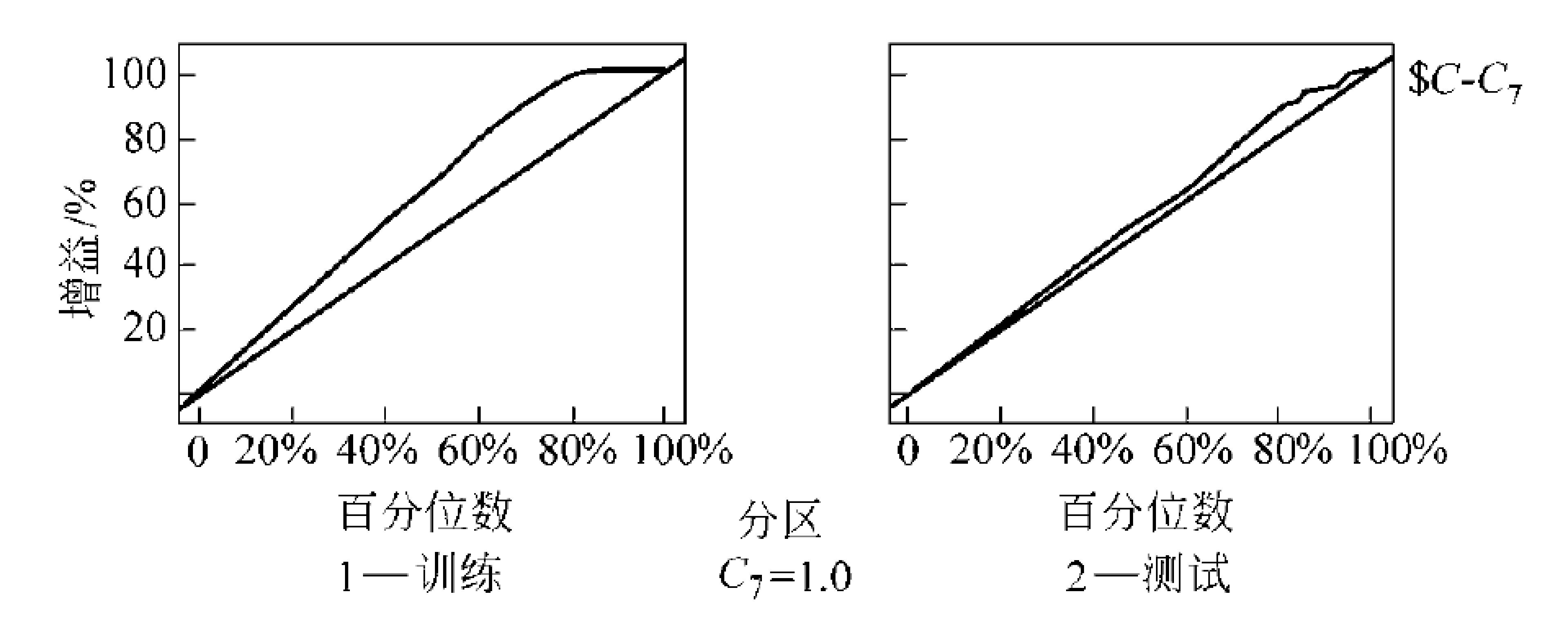


图 12-3　生活满意度的思维方式分析决策树模型增益图

提高难因之职业教育落后、认为子女应接受教育的最高程度。

对教育环境满意的 88% 的村民对生活都较满意，而对教育环境不太满意的对生活满意度不定。希望子女受教育在教学水平上需改进的，多对生活满意度较高；认为子女只要认字就可以的对生活满意度不高；认为子女需接受教育水平为大学的，70% 的村民生活满意度高。

第13章　基于面板模型的农村现代化对农民生活满意度的影响分析

在上一章中，我们运用问卷所收集到的数据进行了基于决策树的数据挖掘分析，数据的来源——问卷代表了农民生活满意度的主观真实反映，从上面部分的分析我们得到了生产生活和个人因素对农民生活满意度的影响机理。但为了从客观方面更充分地解释生产生活因素对农民生活满意度的影响，收集了新农村主要调查村落所在地，即石家庄、邢台和唐山这3个城市相关因素的指标数据，并进行面板数据分析。

面板数据(panel data)又称为平行数据(longitudinal data)，指在时间序列上取多个截面，在这些截面上同时取样本观测值所构成的样本数据。也就是把截面数据和时间序列数据融合在一起的数据。相对只利用截面数据模型和只利用时间序列数据模型进行经济分析而言，面板数据模型具有许多优点：①包含更多的变量，这样就增加了自由度，减少了解释变量之间的共线性，从而改进了估计的有效性；②从多种层次、不同个体的角度进行经济问题分析；③面板数据比单纯的时间序列数据或横截面数据能够更好地识别和测量一些效应；④面板数据通常以微观单元来收集，在微观的水平上许多变量能被更为精确地测量，因此可以有效减少测量误差引起的偏差。

13.1　面板模型变量选取

从上部分分析我们得出结论，农民收入极大地影响着农民生活满意度，故我们在此部分，使用农民收入作为因变量。生产因素部分选择

了机耕面积 agrimech、机械总动力 machinery、各市农林牧渔业总产值 ffaf、各市农村工业总产值 vnonagri、农业产业化龙头经营组织联系农户数 corporation，生活因素部分选择了农村用电量 ele、乡村从业人员人数 worker、各市城乡商品市场成交额 market、农村人口数 population 等指标。在进行检验和建立模型过程中，由于存在多重共线性等原因去除了一些变量。

13.2　面板数据模型的介绍

13.2.1　面板数据的单位根检验和协整检验

1. 单位根检验

对面板数据考虑下面的 AR(1)过程：

$$y_{it}=\rho_i y_{it-1}+x'_{it}\delta_i+u_{it},\quad i=1,2,\cdots,N;t=1,2,\cdots,T_i \tag{13-1}$$

根据对式(13-1)中参数 ρ_i 的不同限制，可以将面板数据的单位根检验方法划分为相同根和不同根两大类。相同根情形下的单位根检验，假设面板数据中的各截面序列具有相同的单位根过程，即假设式(13-1)中的参数 ρ_i 满足 $\rho_i=\rho(i=1,2,\cdots,N)$，包括 LLC 检验、Breitung 检验、Hadri 检验等。不同根情形下的单位根检验方法允许面板数据中的各截面序列具有不同的单位根过程，即允许参数 ρ_i 跨截面变化，包括 Im-pesaran-Skin 检验、Fisher-ADF 检验和 Fisher-PP 检验等。

2. 协整检验

协整概念是一个强有力的概念，因为协整允许刻画两个或多个序列之间的平衡或平稳关系。对于每一个序列单独来说可能是非平稳的，而这些时间序列的线性组合序列却可能有不随时间变化的性质。面板数据的协整检验方法可以分为两大类：一类是建立在 Engle and Granger 二步法检验基础上的面板协整检验，具体方法主要有 Pedroni 检验和 Kao 检验；另一类是建立在 Johansen 协整检验基础上的面板协

整检验。本章将采用 Pedroni 检验法进行分析。

13.2.2 面板数据模型形式

含有 N 个个体成员方程的 Panel Data 模型形式如下：

$$\boldsymbol{y}_i = \boldsymbol{\alpha}_i + \boldsymbol{x}_i\boldsymbol{\beta}_i + \boldsymbol{u}_i, \quad i = 1,2,\cdots,N \tag{13-2}$$

其中，$\boldsymbol{y}_i$ 是 $T\times1$ 维被解释变量向量，$\boldsymbol{x}_i$ 是 $T\times k$ 解释变量矩阵，$\boldsymbol{y}_i$ 和 $\boldsymbol{x}_i$ 的各分量是个体成员的经济指标时间序列，例如，若个体成员代表各不同地区，则 $\boldsymbol{y}_i$ 和 $\boldsymbol{x}_i$ 代表 i 地区指标的时间序列。截距项 $\boldsymbol{\alpha}_i$ 和 $k\times1$ 维系数向量 $\boldsymbol{\beta}_i$，其取值受不同个体的影响，$\boldsymbol{u}_i$ 是 $T\times1$ 维扰动项向量，满足均值为零、方差为 σ_u^2 的假设。

13.2.3 模型形式设定检验

模型(13-2)常用的有如下 3 种情形：

(1) $\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_j$，$\boldsymbol{\beta}_i=\boldsymbol{\beta}_j$(不变系数模型)

(2) $\boldsymbol{\alpha}_i\neq\boldsymbol{\alpha}_j$，$\boldsymbol{\beta}_i=\boldsymbol{\beta}_j$(变截距模型)

(3) $\boldsymbol{\alpha}_i\neq\boldsymbol{\alpha}_j$，$\boldsymbol{\beta}_i\neq\boldsymbol{\beta}_j$(变系数模型)

第(1)种称为不变系数模型，是指在横截面上既无个体影响，也无结构变化；第(2)种情况称为变截距模型，是指个体在横截面上所受影响不同，变截距模型又分为固定影响和随机影响两种情况；情形(3)称为变系数模型，这种情况下的经济结构除了存在个体影响外，在横截面上还存在变化，因而模型的参数在不同横截面上是不同的。在对面板模型进行估计时，如果模型形式设定不正确，估计结果将与所要模拟的经济现实偏离甚远。我们要建立更精确的模型，尽可能地避免误差。因此，建立面板数据模型的第一步便是选择一个最适合的模型，此时需要检验样本数据究竟符合上面 3 种面板数据模型中的哪一种。对上述 3 种模型形式的检验通常使用协方差分析检验法，主要检验如下两个假设：

$$H_1: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_N$$

$$H_2: \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \cdots = \boldsymbol{\alpha}_N, \quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_N$$

该检验通过 F 检验来进行，检验假设的 F 统计量分别为

$$F_1 = \frac{(S_2 - S_1)/[(N-1)k]}{S_1/(NT - N(k+1))} \sim F[(N-1)k, N(T-k-1)] \tag{13-3}$$

$$F_2 = \frac{(S_3 - S_1)/[(N-1)(k+1)]}{S_1/(NT - N(k+1))} \sim F[(N-1)(k+1), N(T-k-1)] \tag{13-4}$$

其中，S_1 为情形(3)(变系数模型)的残差平方和；S_2 为情形(2)(变截距模型)的残差平方和 ；S_3 为情形(1)(不变系数模型)的残差平方和。

首先对原假设 H_2 进行检验，如果所得结果接受假设 H_2 则可以认为样本数据符合情形(1)，即模型为不变参数模型，无需进行进一步的检验。如果拒绝假设 H_2，则需检验假设 H_1。如果接受 H_1，则认为样本数据符合情形(2)，即模型为变截距模型；反之拒绝 H_1，则认为样本数据符合情形(3)，即模型为变参数模型。

根据个体影响的不同形式，变截距模型和变系数模型又分为固定影响模型和随机影响模型两种。在确定模型是固定影响还是随机影响时，可以运用 Hausman 检验法。

13.3　面板数据单位根检验

具有协整关系的面板数据才可以建立面板数据模型，因此在建立面板模型之前，首先要对面板数据进行平稳性检验和协整检验。变量的平稳性是建立时间序列模型的重要前提。沃森(1989)证明当变量存在单位根即非平稳时，传统的统计量(如 T 值、W 值和 R 值等)将出现偏差。Granger 和 Philips 指出，当使用非平稳序列进行回归时，将会造成虚假回归。原假设为序列存在单位根，括号外为统计量的值，括号内为相伴概率，具体检验结果如表 13-1 所示。

表 13-1 面板 ADF 单位根检验结果

变　量	检验形式(C,T,K)	ADF	是否平稳
$\Delta_{income2}$	(C,T,0)	22.1760(0.0011)	是
$\Delta_{agrimech}$	(0,0,0)	26.1971(0.0002)	是
$\Delta_{machinery}$	(C,T,0)	15.7632(0.0151)	是
Δ_{ffaf2}	(C,T,0)	16.1279(0.0456)	是
$\Delta_{corporation}$	(C,0,0)	15.5047(0.0167)	是
Δ_{car}	(C,T,0)	17.9935(0.0062)	是
$\Delta_{population}$	(C,0,0)	13.9540(0.0302)	是
Δ_{ele}	(C,0,0)	13.0046(0.0430)	是
$\Delta_{market2}$	(C,0,0)	19.1129(0.0040)	是
Δ_{worker}	(C,0,0)	23.3509(0.0007)	是

注：Δ 表示序列的一阶差分；检验形式中 C 和 T 分别表示检验方程中带有常数项和趋势项(0 表示没有该项)；之后期数以 AIC 最小准则确定。

13.4　影响农民生活满意度的农业生产因素的面板模型分析

在 13.2 节面板数据模型介绍中已经提到，含有 N 个个体成员方程的面板数据模型有 3 种类型：无个体影响的不变系数模型、变截距模型、含有个体影响的变系数模型。我们统计建模的出发点是试图找出最能反映出各项指标对收入影响的模型，然后借助模型进行实证研究和理论分析。

13.4.1　协整分析

由于面板数据存在单位根，而且均是一阶单整序列，因此可以继续判断几个指标间是否存在协整关系。为保证结论的可靠性，本节采用佩德罗里(Pedroni)提出的 4 个检验统计量判断收入 income、机耕面积 agrimech、各市农林牧渔业总产值 ffaf、各市农村工业总产值 vnonagri、农业产业化龙头经营组织联系农户数 corporation 5 个变量间是否存在协整关系，原假设为变量之间不存在协整关系，括号外为统计量，括

号内为概率值，具体检验结果如表 13-2 所示。

表 13-2　农业生产因素面板数据协整检验

检验方法	检验假设	统计量名	统计量值(P 值)
Kao 检验	$H_0: r=1$	ADF	−2.436 773(0.0074)
Pedroni 检验	$H_0: r=1$ $H_1: (r_i=r)<1$	Panel PP-Statistic	−4.537 096(0.0000)
		Panel ADF-Statistic	−2.908 054(0.0059)
		Group PP-Statistic	−7.255 332(0.0000)
		Group ADF-Statistic	−3.781 606(0.0035)

从表 13-2 中可以看出，4 个统计量都拒绝了变量间不存在协整关系。可以认为在收入 inincome、机耕面积 inagrimech、各市农林牧渔业总产值 inffaf、各市农村工业总产值 invnonagri、农业产业化龙头经营组织联系农户数 incorporation 之间存在长期均衡稳定关系。

13.4.2　影响农民生活满意度的农业生产因素的面板模型选择

首先建立生产因素 3 种形式的模型，得到 S_1 为变系数模型的残差平方和，为 0.183 280；S_3 为不变系数模型的残差平方和，为 1.253 556；N 为截面成员数目，为 3；T 为截面成员的观察时期，为 10；$k=4$。计算得出 $F_2=8.759\ 399$，而临界值 $F(10,18)=2.82$。因此拒绝模型为不变系数模型的原假设，$S_2=0.312\ 448$，计算得出 $F_1=1.321\ 439$，$F(8,18)=3.17$，选择变截距模型。

数据来源于河北省经济统计年鉴和石家庄统计年鉴，实证结果由计量专业软件 EViews 6.0 完成。

13.4.3　影响农民生活满意度的农业生产因素的面板模型建立

对生产因素面板数据引入个体固定影响变截距模型进行讨论，使用截面加权最小二乘回归，估计结果如下：

$$\text{inincome}_{it} = -4.723 + \alpha_i^* + 0.214\text{inagrimech}_{it} + 0.227\text{invnonagri}_{it} + 0.488\text{inffaf}_{it} + 0.013\text{incorporation}_{it}$$

$R^2 = 0.9493$， $F = 307.2076$

其中，α_i^* 为 i 地区的个体固定效应因子对模型得到的平均收入水平的偏离，其中石家庄为－0.102 293，唐山为－0.013 837，邢台为0.116 130。

从模型中我们可以看出农民收入呈现出积极的正面作用。其中农林牧渔业仍旧是农民收入的主要影响因素，其系数达0.487，农林牧渔业总产值是以货币表现的农林牧渔业的全部产品总量和对农林牧渔业生产活动进行的各种支持性服务活动的价值，说明第一产业对农民满意度影响依旧较大。其次，农村工业总产值对农民收入的影响系数为0.227 04，说明农村工业带动了农民收入的增长；机耕面积代表着农村机械化程度，对农民收入的影响也较大。在农业的生产经营过程中，用机械的、电动的各种农业机具，取代传统手工工具、畜力为动力的农具，凡是可以使用机器的地方都使用机器。不单耕作机械化，而且加工、运输、储藏、烘干等生产经营过程全部机械化，以减轻劳动强度，提高效率，降低成本，增加农民收益。农业产业化龙头经营组织联系农户数对农民收入的影响为正，系数较小。这是由于使用的变量为龙头经营组织联系户数，户数变化量较小，不如产量增加幅度大，且联系户数受城镇化农民外流现象的影响。

13.5 影响农民生活满意度的农村生活因素的面板数据分析

13.5.1 协整分析

根据表13-1的单位根检验结果可知，影响农民生活满意度的生活因素的面板数据存在单位根，而且均是一阶单整序列，因此可以继续判断几个指标间是否存在协整关系，协整检验结果见表13-3。

表 13-3　生活因素面板数据协整检验

检验方法	检验假设	统计量名	统计量值(P 值)
Kao 检验	H_0：$r=1$	ADF	－3.936 422(0.0000)
Pedroni 检验	H_0：$r=1$ H_1：$(r_i=r)<1$	Panel PP-Statistic	－8.223 149(0.0000)
		Panel ADF-Statistic	－3.790 462(0.0003)
		Group PP-Statistic	－12.522 55(0.0000)
		Group ADF-Statistic	－2.562 249(0.0150)

从表 13-3 中可以看出，4 个统计量都拒绝了变量间不存在协整关系。可以认为在收入 inincome、乡村从业人员人数 inworker、农村用电量 inele、各市城乡商品市场成交额 inmarket 之间存在长期均衡稳定关系。

13.5.2　影响农民生活满意度的农村生活因素的面板模型选择

建立生活因素 3 种形式的模型，得到 S_1 为变系数模型的残差平方和，为 0.191 175；S_3 为不变系数模型的残差平方和，为 0.909 963；N 为截面成员数目，为 3；T 为截面成员的观察时期，为 10；$k=4$。计算得出 $F_2=5.639\ 76$，而临界值 $F(10,18)=2.82$。因此拒绝模型为不变系数模型的原假设，$S_2=0.230\ 678$，计算得出 $F_1=2.262\ 436$，$F(8,18)=3.17$，选择变截距模型。

变截距模型又分为固定影响变截距模型和随机影响变截距模型。进一步考虑模型是受固定影响还是随机影响。分别建立随机影响模型，进行 Hausman 检验。Hausman 检验的原假设模型为随机影响变截距模型，计算 P 值为 0.0000，得到两个面板模型均为固体影响变截距模型。

13.5.3　影响农民生活满意度的农村生活因素的面板模型建立

对生活因素面板数据引入个体固定影响变截距模型进行讨论，使

用截面加权最小二乘回归，估计结果如下：

$$\text{inincome} = 0.421\,776 + \alpha_i^* + 0.174\,566\text{inele}_{it} - 0.121\,384\text{inworker}_{it} + 0.702\,104\text{inmarket}_{it}$$

$$R^2 = 0.935\,418, \quad F = 78.214\,14$$

其中，α_i^* 为 i 地区的个体固定效应因子对模型得到的平均收入水平的偏离，石家庄为−0.721 655，唐山为 0.024 454，邢台为 0.697 201。

从生活因素影响农民收入分析来看，城乡商品市场成交额与农民收入有较强关联。城乡商品成交额增加，农民收入进行二次分配，拉动了农村消费市场，增加了就业岗位，农村经济市场繁荣，反过来提高了农民收入。农村用电量与农民收入正向相关，用电量的增加，农村电视等设备用电量增加，生活现代化水平提高，体现了农民收入的增加。近年乡村从业人数减少，大量农民外出务工，留守在村中的乡村从业人数或承包了其他务工人员的土地，生产资源充足，进行机械化劳作，收入增加，或在村中龙头经营组织或工厂中劳作，收入也得到提高，造成了乡村从业人数与农民收入成反比的结果。

此外，从全省来看，从 2005 年至 2010 年，县镇中心卫生院数从 620 个增至 641 个，乡镇卫生院床位数从 35 560 张增至 57 097 张，乡镇卫生院机构人员数也有大幅提升，从 42 702 人增加至 53 966 人，说明乡、村两级医疗卫生服务网络建设得到了加强。敬老院工作人员数量从 4652 人增至 8162 人，床位数从 35 827 张增至 103 668 张，年末收养老人数从 22 310 人增至 87 241 人，农村最低保障资金从 5318.4 万元增长至 161 499 万元。5 年来石家庄市乡村人口增长 5%，唐山乡村人口增长 3%，邢台乡村人口增长 2%的背景下，农村的医疗、养老、社会保障 3 方面水平得到了大幅提升，提高了农民生活满意度。

13.6　影响农民生活满意度的农民素质的变量分析

由于缺少从 2000 年至 2010 年各市反映农民素质的变量和教育现状的数据，因此，我们选用 2005 年至 2010 年各市农村文盲、半文盲的人数，大专以上及占农村人口的比例进行研究。具体数据见表 13-4。

表 13-4　各市乡村从业人员中文盲、半文盲人数及占该市乡村人口比例

年份	石家庄市	比　例	唐山市	比　例	邢台市	比　例
2005	66 009	0.010 02	26 006	0.004 995	33 236	0.005 762
2006	63 524	0.009 691	23 751	0.004 557	29 799	0.005 122
2007	60 770	0.009 257	25 131	0.004 804	26 369	0.004 479
2008	58 869	0.008 893	21 156	0.004 007	23 231	0.003 917
2009	56 818	0.008 520	20 347	0.003 810	19 597	0.003 283
2010	51 231	0.007 612	17 882	0.003 332	16 878	0.002 789

表 13-5　各市乡村从业人员中大专及以上文化程度人数及比例

年份	石家庄市	比　例	唐山市	比　例	邢台市	比　例
2005	42 373	0.006 432	23 121	0.004 441	6083	0.001 055
2006	44 656	0.006 813	25 158	0.004 827	8228	0.001 414
2007	47 904	0.007 297	32 758	0.006 263	10 976	0.001 864
2008	51 777	0.007 822	38 842	0.007 357	13 016	0.002 195
2009	63 520	0.009 525	43 789	0.008 201	19 147	0.003 208
2010	74 531	0.011 074	57 385	0.010 693	22 779	0.003 764

由表 13-4 和表 13-5 可以看出，从 2005 年到 2010 年各市农村从业人员中文盲、半文盲的人数都逐渐减少，石家庄乡村从业人员中文盲、半文盲的比例下降了 0.24%，从 2005 年的 1%下降为 0.76%，但仍居三大市文盲率之首。唐山市文盲率在 2005 年为 3 市最低，农民文化水平稍高，同时这个比例 5 年来降幅也较小。邢台市乡村从业文盲率降幅最大，下降了 0.3%，2010 年时文盲率为 3 市最低。

在文盲率降低的同时，可以看到，各市乡村从业人员中大专及以上文化程度的人数及比例也在大幅上升。石家庄市从 0.6%达到了 1%，比率一直位于 3 市之首；唐山市大专及以上文化程度比例变化最大，从 2005 年的 0.4%发展为 2010 年的 1%；邢台市大专及以上文化程度的人群位于 3 市之尾，从 0.1%涨为 0.4%。

从全省来看，农村普通中学学校数从 2005 年 2396 所降至 2010 年 1456 所，普通小学数从 17 638 所降至 11 084 所，这是因为国家实施了撤点并校的政策。中学生数从 2005 年 171.62 万人降至 74.89 万人，

小学生数从325.41万人降至309.8万人，普通中学专任教师数从2005年10.15万人降至6.96万人，普通小学专任教师数从21.83万人降至20.84万人，然而从每位学生享有的教师资源来说，普通小学的资源略有增加，普通中学的资源则增幅较大。此外，乡镇文化站从2005年的1833个增长至2010年的1938个，增加了5.7%，推动了农村文化建设，提高了农民文化素质。

第 14 章　农村现代化对农民生活满意度的影响机理分析

14.1　农业生产因素对农民生活满意度的影响分析

河北省是农业大省，农村经济主要依靠第一产业，农民收入也主要来源于第一产业。因此，无论是第一产业的政策鼓励情况，还是生产基本条件都对农民生活满意度影响较大。结合基于决策树算法和面板模型的结果，可以知道在生产因素中，畜牧业补贴政策实行情况的满意度对农民生活满意度的影响最大。

畜牧业补贴政策覆盖范围广，补贴力度大，惠农效果显著，因此对农民生活满意度影响尤为明显。仅 2012 年中央财政就安排了农机购置补贴 200 亿元，补贴范围继续覆盖所有农牧业县(场)，补贴机具种类涵盖 12 大类 46 个小类 180 个品目。覆盖范围涵盖重大动物疫病强制免疫补劣政策，支持农产品、农超对接，超市从农民企业合作社购进的免税农产品，抵扣增值税进项税额；降低了“农超对接”门槛，严禁超市向合作社收取不合理费用；多方位的税收优惠政策，对农民专业合作社向本社成员销售的农膜、种子、种苗、化肥、农药、农机，免征增值税等；直接用于农林牧业的生产用地，免征城镇土地使用税；2011 年中央财政安排良种补贴 220 亿元，农机具购置增加到 175 亿元；实行畜牧良种政策。从 2005 年开始，国家开始实施畜牧良种补贴政策；扶持“菜篮子”产品标准化生产政策；中央财政投入资金实施畜禽标准化养殖扶持项目，对主产省区蛋鸡、肉鸡、肉牛和肉羊规模养殖场，支持标准化生产改造。补贴范围全面涵盖农村生产，降低了生产成本，拓宽了农产品销路，使农民生产积极性提高，农民收入提高，生活满意度也相应提高，而

若补贴落实不到位，将对农民产生很大消极影响，生活满意度也将降低。

河北省自然条件优越，种植业生产历史悠久，是我国重要的农业大省，种植业情况对农民生活满意度影响也较大。种植业是国民经济的基础，关系着国民经济的健康发展，而且与人民生活息息相关。2004 年初党中央下发“一号文件”，明确要求“集中力量支持粮食主产区发展粮食产业，促进种粮农民增加收入”，河北省陆续出台了减免农业税及粮食直补、良种补贴、农机补贴、加大财政农业投入等一系列惠农措施，给种植业带来了新的生机。在政策驱动、市场拉动和工作推动的综合作用下，种植业生产再谱新篇章，粮食生产出现了重要转机，经济作物产量继续稳步增长，提高了农民生活满意度。种植业布局受自然条件、社会经济条件和传统种植观念的影响，反映了农业生产基本态势，对农业生产力的提高具有影响作用，并直接影响粮食安全及区域经济可持续发展。然而随着土壤、灌溉、生产投入、交通、社会经济等条件的改善，种植业传统布局若不够合理，则会制约农业经济的发展，影响农民生活满意度。

农业机械化是农业现代化的中心环节，它凝聚着现代科学技术的最新成果，并配合农业生物等农业技术，成为发挥增产作用的基本手段和提高劳动生产率、减轻繁重体力劳动的必要条件和根本途径，从而带来生产力的质的飞跃。农业机械化作业领域由粮食作物向经济作物，由大田农业向设施农业，由种植业向养殖业、农产品加工业发展，由产中向产前、产后延伸，发展空间不断扩大。在农业的生产经营过程中，用机械的、电动的各种农业机具，取代传统手工工具、畜力为动力的农具，凡是可以使用机器的地方都使用机器。不单耕作机械化，而且加工、运输、储藏、烘干等生产经营过程全部机械化，以减轻劳动强度，提高效率，降低成本，增加农民收益，农业机械化水平的提高带动了农民生活满意度的提高。

14.2　农村生活因素对农民生活满意度的影响分析

结合决策树算法和面板模型的结果可知，对农民生活满意度来说，生活因素中，农村基础设施建设和城乡商品市场的情况的影响力最强。近几年来农村基础设施建设明显加强。良好的基础设施和公共服务是农村经济增长的重要基础和源泉，也日益成为农村持续发展的内在变量。以水、路、气、电、信为重点的农村基础设施建设加快推进，在基本实现村村通油路或水泥路的基础上，推进农村客运网络化建设和线路公交化改造，基本实现了具备通车条件的行政村通客车的目标；大力实施农村电网改造，农村基本实现户户通电；50 户以上的自然村实现了通广播电视；全省新解决了 366 万农村人口的饮水安全问题；新建农村沼气 16 万户，累计完成 300 多万户，总量居全国第四位。饮用水、公共道路、通信设施等基础设施的完善对农民生活是最现实的影响，因此对农民生活满意度影响最高。城乡商品市场的繁荣会给农民生活带来便利，消除因距离带来的物价差距，此外，城乡商品市场的繁荣又可以解决附近村民的就业，对农民生活满意度影响也较高。

其次，农村生活因素中社会保障制度满意度对农民生活满意度的影响较强，包括最低保障制度、医疗保障和养老保障。以市场为导向的农村改革，使经济快速发展，但同时也扩大了收入分配差距。市场经济的发展情形下，公平是不能自发实现的，更不能自动地保障弱者的生存和发展。尤其是近年来，农业增效受市场的制约难度越来越大，农民收入增长势头放缓，增收的不均衡性、波动性、困难性特征更加明显，农村部分低收入者的生活受到了比较大的影响，部分灾民和贫困户的生活困难问题更为突出，建立健全农村社会保障体系显得尤其迫切。医疗养老等问题是农民收入中占比最高的两个项目，新型农村合作医疗和养老保险的逐步实施，落户农村，将给农民减轻极大的负担。因此，社会保障的落实情况影响到这些弱势农民群体的切实生活，而若在此环节产生权力寻租等不公平现象，将对整个农村社会产生恶劣影响，影响社会公平和政府公信力。故此社会保障的落实对农民生活满意度的影

响也较大。

然后，收入来源对生活满意度的重要性位于第三位。随着农村工业化、农业产业化和乡镇企业的发展，农村就业结构发生变化，收入分配的不均也逐渐体现在农村中，农业产业化经营以市场为导向，以家庭承包经营为基础，依靠各类龙头企业和组织的带动，将生产、加工、销售紧密结合起来，实行一体化经营。这种经营方式，改变了只从事原料生产的传统农业，提高了农业的整体效益，增强了农产品的国际竞争力。特别是龙头企业一头连着市场，一头连着农户，它可以把市场信息、适用技术、管理经验传送给农户，给农民带来巨大收益，而仅凭家庭自主劳作的收入较低，因此村民收入来源对农民生活满意度的影响变得越来越大。

14.3　农民素质对农民生活满意度的影响分析

农民素质的提高会容易获得自我肯定和精神提升，会获得尊严感、认同感、归属感和精神愉悦。因此，农民素质对农民生活满意度意义重大。而其中，村民对教育环境的满意度、教育条件的满意度，对提高农民素质培训的满意度，对农民生活满意度影响较大。

农民作为特定的社会阶层和社会劳动群体，其素质符合人口素质的一般构成，但又有其自身的特点。本章讲的农民素质，是从劳动者素质的角度来谈的，是指在农村这一特定环境的影响下，从事农业生产所应具备的基本自身条件和基本能力，包括农业劳动者的体质、品质和素养等。在社会主义新农村建设实践中，从事农业专业化生产和产业化经营的新型农民应该具备的基本素质同样是由多个方面构成的。2006 年中央一号文件《关于推进社会主义新农村建设的若干意见》中明确提出：“提高农民整体素质，培养造就有文化、懂技术、会经营的新型农民，是建设社会主义新农村的迫切需要。”也就是说，“有文化、懂技术、会经营”应当是现代新型农民的基本内涵。

教育是发展科学技术和培养人才的基础。钱克明的研究表明：“每增加 1 元农村教育投资，可使农牧业产值增加 8.43 元，而每增加

1 元农村公共基础设施投资,仅可使农牧业产值增加 6.75 元。且政府每增加 1 元的农村教育投入,可减少农牧户物质费用 6.64 元。”教育投资还能提高工资水平,“每增加 1 年的教育,农户平均工资性收入水平至少可以增加 1000 元。”由此可见,在建设社会主义新农村中,必须把开展对农民教育文化的工作摆在最优先的战略地位。随着农民越来越渴望自身素质提升,他们也越来越关注子女的教育问题,因此子女教育环境和条件是农民生活满意度的影响因素。

目前,我国农业正在由传统农业向现代农业转型。建设现代农业需要用先进的物质条件装备农业,用先进的科学技术改造农业,用先进的组织形式经营农业,用先进的发展理念指导农业。通过加快农业科技进步,加强基础设施建设,调整农业产业结构,转变农业发展方式,提高农业综合生产能力。在农业生产经营过程中,先进的生产工具、科学技术、管理经验以及经营体制和运行机制要靠人去探索、去应用,无论是发展方式的转变、产业化链条的建立,还是生产绩效的提高,都需要在人的主观能动作用下得以实现。农民是农业市场经营的主体,是农村生产力中最活跃的因素,农业和农村经济的发展,最终取决于农民积极性和创造性的发挥,农村现代化的实现最终取决于农民的现代化。只有使我国农民的生产技能素质由掌握传统生产技能向掌握现代生产技能转变,才能推动“生产发展”。提升我国农民的素质,是推动“生产发展”必不可缺的条件。农民也逐渐意识到知识和技术能够使农业生产更加方便快捷,提高生产效率,增加收入,对培训的渴望也就越来越真切,因此对农民技术上的培训是影响农民生活满意度的又一重要方面。

第 15 章 “提升农民生活水平，实现‘三农’协调发展”的对策建议

15.1 河北省农村能源发展对策

通过对河北省农民能源使用状况总体特征的研究，我们找到了河北省农民能源使用过程中存在的各种问题。再通过因子分析不仅让我们看到了河北省地区间农民能源的消费差异，也让我们看到了地区农民能源消费差异的大小，更让我们找到了消费差异的原因。通过差异状况找到原因，再通过原因探索符合河北省情及各市市情的解决问题策略是我们研究此课题的又一方法论指导。在这一方法论的指导下，我们认为研究河北省农民能源发展战略，改善河北省农民能源使用状况意义重大，并可以从下列几个方面着力。

1. 从实际情况出发，着力提高农民收入水平

随着“三农”问题的不断解决，新农村建设的不断深入，河北省农业和农村能源建设也将不断面临新的挑战。而在这些挑战中最突出的问题是如何千方百计增加农民的收入。只有农民收入增加了，才有能力改善能源的消费结构，才有能力使用污染低、效率高的能源，才有能力改善农村能源的基础设施建设状况。河北省是一个农业大省，地处华北平原，环抱京津，东临渤海。但是，2006 年，全省农民人均收入仅为 3801.82 元，居全国第十，排东部沿海大省之末。并且全省 11 个市中，农民人均收入大于 4000 元的仅有唐山(5155 元)、廊坊(4950 元)和石家庄(4486 元)。其余 8 个市农民人均收入均在 4000 元以下，尤其张家口(2670 元)、承德(2670 元)农民人均收入分别低于全省平均水平达 1131 元和 951 元。由实证过程可以深刻认识到，收入水平高的唐山、

廊坊、石家庄各项与经济社会发展正相关的指标得分均处全省前列。而张家口、承德因收入低，农民能源无论是生活用能状况（综合排名分别为第十、第八）还是生产用能状况（综合排名分别为第十一、第十）基本均处于全省劣势地位。因此农民增收应当是河北省当前农村能源建设的核心问题。

2. 提高农村煤炭的利用效率

随着河北省经济社会发展水平和文明程度的不断提高，以及农村居民收入的不断增加，再加上河北省具有开滦集团精煤生产基地、邯郸炼焦煤和动力煤基地、蔚州矿区动力煤基地等煤炭基地的供煤保障，农村居民生活用能弃薪柴而代之以煤炭及煤炭的转化能源是必然趋势。从普查数据来看，全省农村共有 611.75 万户主要炊事用能为煤，另外还有 252.76 万户辅助炊事用能为煤。其中邯郸市主要炊事用能为煤的农户比重达 93.28％，石家庄市达 83.87％，邢台市达 76.98％，保定市达 53.13％。

这说明煤在农民日常生活能源中已占据了相当重要的位置。但是目前全省仍有 48.72％的农户主要生活炊事用能为柴草，尤其承德市该比重达 90.84％，沧州市达 89.74％，秦皇岛市达 78.73％，衡水市达 77.56％，唐山市达 74.24％。这说明煤炭取代柴草成为主要的生活能源任重道远。因此优化全省农村煤炭供应链、加大煤矿勘探力度、挖掘资源潜力，开发新的接替区域并指导农民提高煤炭的利用效率，减少二氧化碳的排放量，将会对全省农民能源消费结构、环境保护产生深远的影响。

3. 加快全省农村的电源、电网建设

随着收入水平和物质文化水平的提高，农民对电视、电扇、冰箱、洗衣机等耐用电器消费品的需求越来越大。尽管目前全省平均通电的村比重达 99.5 个百分点，但各地区通电情况存在差异，尤其唐山、邯郸等市均多达 51 个村没有通电。石家庄也有 34 个村没有通上电。因此加快全省农村的电源、电网建设，优化电源布局，以最短的路线满足最多农村的用电需求将不仅直接关系到农民生活质量的提高，更关系到农

村能源消费结构的合理布局。

4. 大力发展农村沼气产业，建设节约型新农村

尽管河北省农村目前有6.01万户主炊事能源使用沼气，另有17.76万户辅助炊事能源使用沼气，但全省目前利用柴草作为主炊事能源的农户数量达669.94万户，是沼气用户的111.42倍。这说明全省仍需要大力发展农村沼气产业，减少对污染高、效率低的柴草的使用。从沼气的运作机理来看：种植业中的秸秆可过腹转化为牲畜粪便或直接入池发酵，产生的沼气可用于农户炊事或照明，也可发电用于农产品加工业，从而减少了外界能源的输入。沼渣可作为农业生产的底肥，也可以与氮、磷、钾等一起加工成颗粒肥，还能作为培养基用来生产食用菌和养殖蚯蚓。沼液和沼渣中含有多种氨基酸、生长素、抗生素和多种微量元素，沼液直接用于叶面施肥，既能增加作物产量又可以提高品质，增强植物的抗逆性。沼肥能改良土壤，增加土壤的团粒结构，增强土壤的保水、保肥、保温和通透性，同时改善土壤的微生态环境，对提高土地的持续生产力具有重要的作用；同时减少化肥和农药的施用量，使农产品成为无公害的绿色食品和有机食品，增强了食品的安全性和在国际市场上的竞争力。因此，沼气的多功能生态效益为农业增效和农民增收、为农业的可持续发展奠定了坚实的基础。2006年，河北省主要农作物总播种面积达877.737万hm^2，粮食总产量达2702.8万t，这其中反映了农村可供作沼气用的秸秆资源非常丰富。因此根据各地区实际情况，在已有沼气规模的基础上，进一步加强对农民关于沼气产业的宣传、培训技术支持，尤其是完善优惠政策，增加资金投入，抓好典型沼气试验区的建设，加快推广农村户用及大中型沼气工程将大大改善农村能源消费结构，提高生态效益和经济效益，并使化肥和农药的用量降低，对美化农村环境、提高农民的生活质量和精神文明水平也将具有重要的意义。

5. 全面发展农村可再生新能源，实现农村多能互补的能源模式

农村能源建设是一个系统工程，应以满足当地农村建设为目标，以各地区的本地资源为基础建立能源系统，实现“多能互补，综合利用”。

综观河北省 11 个地区的实际情况,全省水力、风力、地热、太阳能、生物质能等丰富。其中全省水能理论蕴藏量 302 万 kW,可开发量 156 万 kW;在张家口、承德坝上地区的风能可装机容量 1370 万 kW;衡水市特有的地层构造为地热资源的形成提供了良好的热储条件,全市平均地温梯度值 3.16℃/100m,略高于大地梯度背景值(3℃/100m),垂向上可分为中低温热水、高矿化热水和基岩高温热水三个热储层。另外廊坊辖区内各区、市、县均有地下热水分布,总面积达 1007.9km^2,出口最高温度达 93℃,极具开发价值。这些得天独厚的条件为河北省农村实现“多能互补,综合利用”的能源模式打下了良好的基础。如何建立一个各种能源有机联系、多能互补,以最低成本获得综合利用效益的集成系统便成为摆在全省各级政府、专家学者面前的难题。

6. 依靠政府指导和政策调节,逐步改善农村能源的使用状况

首先,完善能源价格机制。发挥价格杠杆在农民能源使用过程中的调节作用,加快建立反映农村能源供求状况、农民能源消费结构并能体现各地区能源稀缺程度、节能降耗的价格机制。制定发展农村清洁能源和可再生能源价格鼓励政策,完善对高耗能行业的差别价格政策和峰谷分时电价政策。逐步理顺能源比价关系。其次,加大财税政策对农村能源基础设施建设的支持力度。加大农村节能资金投入,落实国家对节能和可再生能源、新能源发展的财税优惠政策,加大公共财政对节约资源和节能改造的支持力度。对财政性资金支持的建设项目,优先使用节能产品、节能技术、节能材料。还有,在农村加快推进能源法制建设。大力宣传和贯彻落实《中华人民共和国煤炭法》、《中华人民共和国电力法》、《中华人民共和国可再生能源法》等相关法律法规,制定实施细则,坚持依法行政、严格执法,为农村能源发展提供法律保障。

15.2 河北省新农村(民居)建设建议

1. 加大力度进行新农村建设和新民居建设,促使农村产业发展,增加农民收入

党的十六大报告指出:“农村富余劳动力向非农产业和城镇转移,

是工业化和现代化的必然趋势。”在我国，有限耕地上农业劳动力的总量居高不下，是造成农业生产率低、农民收入难以增加的根本原因，而农民增收所面临的最大困难，也恰恰在于农业过剩劳动力向二、三产业和城镇转移困难。这句话其实为我们提供了两个解决方向：一是促进农村劳动力向城镇转移，农民进城务工已成为转移农村劳动力快捷有效的一个现实出路。它对增加农民收入、繁荣城市经济发挥了积极的作用，但是从某些方面，农村富余劳动力的无序流动也会给社会带来负面影响；二是加大对新农村建设投资，促进农村相关产业发展，同时在农民能接受的范围内进行村庄重组改建，落实新民居建设，使农民居住相对集中，同时建设完善医院、学校等公共设施，从而促使农村居民中不愿外出的剩余劳动力有用武之地，最大限度地提高农村劳动生产率。

2. 在新农村建设和新民居建设中做到统筹兼顾、因地制宜

以往的学者研究表明，同样模式的新农村建设和城镇化发展过程中往往存在着不小的差异。我们在调研中发现，新民居建设模式相似的村庄，发展水平往往存在着差异。以唐山市李各庄村和其邻近小区银联花苑（由邻近几个村庄居民和外来务工者为成员的居民小区）为例，其建设模式均是政府同意建造的住宅区，但是通过调查发现，李各庄村居民对于生产、生活各方面均比较满意，但是其邻近的银联花苑居民对于很多方面满意程度很低。究其原因，李各庄村居民多数在唐山市工作，务农村民比例较少，而银联花苑居民多数为务农、做生意，需要一定的存物、存粮空间，而两村居民均为单元楼式建筑，故银联花苑布局显得不合理，存在乱摆乱放现象。因此，政府在统一建造的同时，一方面要针对村庄环境，尽量使村民集中发展；另一方面，要考虑到村庄独有特征，制定相关的建设模式，避免采用单一模式建造。

3. 加大科技投入，促进农民素质提高

培育新型农民的措施中，完善教育体系是基础环节。面对我国农民受教育程度普遍偏低，所掌握的科技知识相对有限的现实，应在大力发展面向青少年的基础教育和职业教育的同时完善面向成人的成人教育。党的十七届三中全会对大力办好农村教育事业做出全面的部署，

并首次提出“重点加快发展农村中等职业教育并逐步实行免费”,这是继全面实施免费义务教育之后,党和国家发展农村教育、促进教育公平、提高农民科学文化素质、缩小城乡差距的又一重大举措。通过对新农村建设中紧缺人才类型的分析,鼓励高校设置相关的农业专业,做到农业专业化人才的培养与新农村建设相结合。

4. 政府主导,促使农村居民提高政策理解度

在新农村民居建设中,政府在规划、建设上应该很好地发挥其主导作用,一方面,要加大宣传力度,使农民加深对新农村建设和新民居建设的理解;另一方面,在农村经济还比较脆弱的地区和民居建设初期的地区,政府在做好前期规划、设计、动员、引导,做好服务等工作的同时,重点放在资金扶持上,不能只追求居住形式上的集中,还要加大力度增加后续建设,尤其大力发展农村科技文教事业,解决农民当前最为关注的问题,满足农村居民最迫切的要求,各方面统筹兼顾,协调发展。

15.3 提高农村现代化水平,提升农民生活满意度的对策建议

15.3.1 推进农业生产现代化 提高农民生活满意度

1. 推进农村各项事业的发展,促进农村繁荣和发展方式转变

为了提高农民生活满意度,需继续推进农村各项事业的发展,促进农村繁荣,为新农村建设提供强有力的产业支撑,推进农村二、三产业结构调整和发展方式转变。实施主导产业推进战略,加快农产品加工业发展步伐。实施传统产业振兴战略,加快产业优化升级。以产业发展为支撑,以强农惠农政策为保障,充分挖掘农业内部增收潜力,巩固提高农民生产经营性收入,着力增加农民工资性收入,不断提高农民转移性收入和财产性收入,确保农民收入持续稳定增长。

2. 加快变革传统农业的生产方式,使之产业化和现代化

在生产条件和方式上,必须以科学技术成果推进农业基础设备的改进,把科技进步转化为农业生产力,使依靠消耗自然资源为主,以“粗

放型”增长方式发展的传统农业，向依靠科技投入为主的“集约型”增长方式过渡，最终形成优质高效、可持续发展的新型现代农业。

在生产技术上，生物工程等高新技术的发展给农业带来前所未有的机遇，必须推进农作物生长的自然规律和种植技巧等知识的创新及在农业生产中的普及，农业的科技贡献率提高，可以促使传统农业由依靠自然规律向发现和改造自然规律的新兴农业转变。

在生产模式上，农业生产过程中自然经济对现代生产力的制约越来越明显，商品经济和市场经济模式也适用于农业市场。注重社会分工之间的协作关系，摆脱自给自足的小规模生产，顺应生产力发展方向，推进农业现代化进程，使农业生产规模化、专业化、区域化。

只有这样，才能逐步改善经济状况，提高农民生活水平，改变贫困面貌。

3. 调整农业产业结构，实现农、林、牧、渔多层次发展

在农业产业结构方面，传统农业经济结构以粮食生产为主，必须改变这种单一的结构，实现农、林、牧、渔等方面多层次发展。如根据地域特色，发展以粮食为主、特产作物(如茶叶)以及饲料种植为辅的多样化种植业；改善以采伐为主的传统林业，逐渐形成以林木培养、林产品提供、提升当地经济发展的环境友好程度三大方面为主要功能的新型产业；改变畜牧业的副业地位，使其成为各类养殖业分化明确的主要产业部门。

15.3.2 推进农村生活现代化　提高农民生活满意度

1. 需加大对农村基础设施建设的投入和支持农业基础设施建设

加大资金支持力度，加强服务和引导，把农村基础设施建设作为调整农业结构、发展农村经济、增加农民收入的“基础工程”来抓，实现农村经济社会的可持续发展。实行农村基础设施经营管理体制改革。对国家和集体投资的基础设施，在确保安全、有效运行，发挥效能的前提下，引入竞争机制，引导当地业主或农民承包经营，实行市场化管理和商业化运作，以缓解设施在运行中的各种经费的不足。对村级一些小型基础设施可引导组建各种合作经济、股份经济、股份合作经济等形

式，实行自负盈亏，自主经营。通过走市场化道路，不仅解决了群众的饮水难题，而且还减轻了群众负担，企业得到了发展，农民生活满意度得到提高。

以农田水利为重点加强农业基础设施建设，在河北省水利资源不够充足的背景下，突出抓好高效节水灌溉建设。改善山丘区农业灌溉条件。全面推进南水北调配套工程建设。随着农村现代化，农村用电量也突飞猛涨，因此需高质量完成新一轮农村电网改造升级工程，切实保障农村居民生活用电，基本解决农业生产用电问题。继续加强农村道路交通建设，不但做好拓宽新的道路，还要加强对旧有道路的保养和维护工作，加大公路通达深度，提高公路网密度。加快实施农村沼气工程，着力推进大中型沼气和集中供气工程建设。

2. 注重分配效率，健全社会保障体系，为经济发展和社会稳定提供良好的环境

在初次分配时注重效率，通过政策干预加强政府对收入再分配的调节职能，建立健全社会保障体系，为经济发展和社会稳定提供良好的环境。医疗养老等问题是农民的重大开支之二，新型农村合作医疗和养老保险的逐步实施，落户农村，将给农民减轻极大的负担。在新型农村医疗保障制度走向 95％覆盖的情况下，养老保险制度逐步完善，可以有效解决农民因病返贫等情况的发生。政府各项补贴政策满意度对农民生活满意度影响也较大。

全面提高农村社会保障水平，包括医疗、养老、低保等社会保障工作。积极在试点地区推进新型农村社会养老保险工作，尽早实现新型农村社会养老保险全覆盖。完善农村五保供养制度，提高供养水平，使其不低于当地居民平均生活水平，帮助困难家庭解决生产、生活中的实际困难。健全农民工社会保障制度，研究制定保障农民基本权益的办法，如养老保险制度的完善，扩大医疗保险覆盖险种。土地是农民生存的根本，因此要对失地农民权益给予充分保障，切实做到先保后征，为失地农民做好长远规划，帮助解决他们的基本就业和未来养老问题。

3. 大力发展农村卫生事业

加强农村社区卫生服务站建设，积极开展以疾病预防、医疗、保健、

健康教育、计划生育等服务为主要内容的社区医疗服务。提高农民健康保障水平，解决农民“因病致贫、因病返贫”问题。在新型农村合作医疗制度全覆盖的基础上，逐步加强财政投入力度，实现门诊统筹全覆盖，让乡镇卫生院积极走开门办院的道路，延伸服务范围。完善多级卫生医疗机构，尤其多设立社区卫生分支机构，方便农民进行问诊，让小病不出村，并建立健康档案，开展健康教育，提供健康咨询等服务。全面提高农村卫生医疗水平。

4. 加强农村环境综合整治，妥善处理农村生活污水和垃圾

地方政府要统筹规划，加快城镇污水处理厂建设和城乡污水管网建设。将有条件的农村地区的污水纳入下游城镇污水处理系统，减少污水对环境的影响。大力开展农村环境卫生综合整治，加快改善农村人居环境。加强畜禽污染治理，制定畜禽养殖污染防治规划，合理确定畜禽养殖规模。严格防控农村工业污染。严格控制高耗能、高污染、资源性和产能过剩项目在农村地区建设。传播农村环境科普知识，增强广大农民的环境意识。保护农村生态环境。坚持农村生态保护与治理并重，加强对矿产、旅游、林业等资源开发活动和交通、水利等生活污染治理基础设施建设的环境管理，遏制生态破坏。深入开展村镇污水处理和城乡一体垃圾处理试点工作。建立健全村庄设施管护、环境保洁、村庄绿化和村容美化等方面的管理制度。

15.3.3 推进农民素质现代化　提高农民生活满意度

1. 要加强农民思想宣传教育

目前仍有许多农民持有“读书无用论”，或认为职业教育“低等”等落后思想存在，导致义务教育即便不收取费用，对职业教育实行多项鼓励政策也无法推行农民文化素质的提高。因此要加强农村思想宣传教育，以村中广播、拉动条幅、义务发放宣传手册、定期在村中活动室举办职业教育宣传活动等方式进行宣传，不断降低农民文盲率，使农民接触文化知识，提高农民文化水平，为更好地掌握农业技术和实现农村文明打下基础。

2．要继续普及九年制义务教育和农村职业教育

建立农村义务教育经费保障机制，对于农村特困家庭儿童给予生活费用补贴，避免其因贫困而放弃义务教育的情况发生。大力改善农村办学条件，加快学校危房改造，实施中小学校舍安全工程，坚决杜绝质量问题，积极推进农村中小学标准化建设；完善农村中小学寄宿制度，保障儿童身心健康发展，在此基础上不断提高农村教育教学水平，尽量避免家长不得不安排村中儿童外出就读的现象发生。加快发展面向农村的职业教育，对中职教育中的涉农专业实行免费政策，定期将教辅教材送至各乡，定期选派涉农教师或科技技术模范农民到各乡进行宣传教育。提高本土农民的素质，培养一批新农村建设带头人。

3．大力实施科教兴农战略

通过多种形式对广大农民进行科普教育，从普及基本知识入手，要紧密结合农村实际情况，当地生产种植情况，有针对性地组织广大农民和农村基层干部学习先进实用的种植、养殖等技术，形成长效培训机制，定点专人专项地对某村进行教育，根据时节性变化，针对可能会在生产时间过程中出现的问题及时与当地养殖户沟通，提高农民收益。除农业生产外，还可以组织普及经营管理、市场营销、卫生保健和法律等方面的基本知识，培训其他相关技能，为到乡镇企业或农业产业化龙头企业工作提供机会。

4．实施农村基层公共文化服务体系建设工程

开展多种形式的农民群众文化活动，精神文化生活。加强农村文化市场建设，营造健康的农村文化环境。大力加强乡镇文化站建设，积极开展“送文化”活动，满足农民群众精神文化需求。繁荣农村文化事业。增加对农村文化事业的投入，加强文化站、文化室等公共文化设施建设，鼓励和支持农民开展业余文化体育活动，扶持发展具有特色的民间民俗文化，使农民的精神境界在健康有益的文化活动中不断得到提升。各级党委、政府要将送文化下乡列入年度经济和社会发展计划，指标纳入考核范围。政府设立专项基金，用于补助各级剧团下乡演出，保证让广大农民多看戏，看好戏。文化单位可加强与科协的合作，在每部

文艺片前放映农业科技宣传片，既满足了农民的需求，又可以集中推广先进的农业科技知识。

5. **繁荣农村文化，形成村中倡导健康文明的新风尚**

加快推进村级“农家书屋”建设，深入推进社会主义文化价值体系，加强公民道德教育、弘扬社会主义道德；加强家庭美德、个人品德和法制教育，发扬中华民族传统美德，推进构建社会主义和谐社会，广泛开展文明家庭、文明村组等文明评比活动，增强农民的公德意识，诚信意识、仁爱意识，引导农民崇尚科学、抵制迷信、破除陋习，推进农村文明化进程，形成文明和谐的新型农村社会氛围。

第 4 篇

适应人口老龄化需求
加快农村养老模式改革

第 16 章 适应人口老龄化需求，开展农村养老满意度调查

16.1 人口老龄化引发的思考

1999 年以来，中国正式步入老龄化国家的行列，人口老龄化问题是我国经济社会发展面临的重大现实问题之一。根据 2010 年第六次全国人口普查的数据，“全国总人口为 133 972 万人，60 岁及以上人口为 17 765 万人，占总人口的 13.26%，其中 65 岁及以上人口为 11 883 万人，占总人口的 8.87%。同 2000 年第五次全国人口普查相比，60 岁及以上人口的比重上升 2.93 个百分点，65 岁及以上人口的比重上升 1.91 个百分点。”预计 2015 年老年人口将超过 2 亿，2030 年将超过 3 亿，2040 年将突破 4 亿，老年人的数量位居世界首位。为积极应对人口老龄化，加快发展老龄事业，根据《中华人民共和国国民经济和社会发展第十二个五年规划纲要》、《中华人民共和国老年人权益保障法》和《中共中央国务院关于加强老龄工作的决定》(中发[2000]13 号)，我国制定了“中国老龄事业发展‘十二五’规划”，以实现老年人“老有所养、老有所医、老有所教、老有所学、老有所为、老有所乐”为工作目标，让广大老年人共享改革发展成果。

我国老龄化社会不同于其他老龄化国家：一是我国老年人口的绝对数大且增长快，老龄化发展迅速；二是地区与城乡之间差异大；三是人口老龄化与严格控制人口增长同步进行；四是与世界上其他发达国家进入老龄化社会时的经济发展水平相比，我国老龄化社会呈现“未富先老”的态势；五是人口老龄化是在社会保障制度还不完善而家庭功能弱化的条件下到来的。

伴随着老年人口的不断增加，独生子女也在不断增加，家庭结构发生变化。子女赡养条件和能力不足，社会负担日益加重。当前我国养老存在着一些问题，一是老龄服务机构尚未建立健全。在养老服务方面，还缺乏一系列基础性的制度安排，受人员、服务、管理等方面的制约，老年事业发展不快，形不成规模。二是老年服务网络不健全。社区服务设施简陋，社会公益设施落后，日常生活中老年人遇到的就医、求助等方面困难不能及时有效解决。三是集中性养老机构规模较小。目前，市、县两区敬养院、老年公寓的建设规模普遍较小，硬件设施条件较差，管理和服务不到位，不能满足日益增长的社会需求。四是我国农村老年人口占比大，由于城乡发展的不平衡，农村经济发展相对落后，农村社会养老保障制度不完善、农民收入水平低下，大量劳动力外流，使得在对老年人生活照料、精神慰藉和疾病护理等方面，都远不及传统的家庭保障，造成农村老人在养老、医疗、生活保障等方面面临更多的困难和问题，农村老龄事业的发展面临更多的挑战。因此，探讨全社会各种因素对农村老年人的生活的影响，探索研究适合农村的养老模式显得尤其迫切。

16.2 农村养老满意度问卷设计与调查

我们以河北省农村养老满意度为切入点，在研究大量资料、参考论文相关实践成果的基础上，以“中国老龄事业发展‘十二五’规划”为指导，分析影响农村老年人养老满意度的因素，综合考虑老年人自身因素、家庭因素、社区因素和国家因素四个方面，设计问卷，在调查研究的基础上，收集实际调查数据，在数据分析的基础上探索适合河北省农村的养老模式，为缓解我们老龄化社会的压力，实现广大农村老年人“老有所养、老有所医、老有所教、老有所学、老有所为、老有所乐”的幸福晚年生活提供有意义的决策支持。

1. 确定调查目的

中国步入老龄化社会，农村老年人口占全国老年人口的比重很高，而且农村养老相较于城镇养老存在更多的困难与问题。农村老年人的

养老问题受到社会各界的关注，本次调查的目的在于全面了解河北省农村老年人养老满意度的整体状况以及影响老年人养老满意度的各种因素，希望为实现“中国老龄事业发展‘十二五’规划”的“老有所养、老有所医、老有所教、老有所学、老有所为、老有所乐”的具体目标，确保老年人生活得更加幸福、更有尊严，促进社会和谐健康发展提供有意义的研究资料。

2．调查方式和调查对象

本次调查的对象是河北省 11 个地市的农村老年人，采用电访调查和面访调查相结合的方式进行。电话调查法是由调查人员通过电话向被调查者询问了解有关问题的一种调查方法，包括传统的电话调查方法和计算机辅助电话调查方法（CATI）。本次调查采用的是计算机辅助电话调查方法（CATI），这种方法最大的优点在于调查过程中质量的监控保证及操作的系统规范化；另外还有覆盖面广和回答率高的特点，任何有电话的地区、单位和个人都可以被访问到，且不受调查者在场的心理压力影响，可以更真实地表达自己的想法。

尽管电访调查有覆盖面广等优点，但是由于农村老年人中手机用户的比例较低，通过电访调查不足以获取足够的样本量，因此，本次调查利用河北经贸大学河北省籍本科生暑假回家的便利条件，组成本科生暑期实习小分队，深入农村社区进行实地的面访调查。由于河北经贸大学河北省的生源较多，且分布广泛，因此此次调查收集到的问卷覆盖到了河北省 11 个地市的农村地区，在实际的调查过程中也兼顾了各市比例的分配。

3．抽样方式及样本量的确定

电访调查利用已有的河北省各市的手机号段，采用随机拨号法（RDD）抽取样本，同时兼顾各市的人口比例确定各市的样本数量。面访调查主要是根据学生的生源地确定问卷发放量，在调查过程中采用的是方便抽样，在街道或居民小区开展调查。方便调查具有一定的局

限性，但由于本次调查是探索性研究，再结合电访调查，因此本次得到的样本也能够达到研究所需要的代表性。

4. 问卷设计与指标选取

老年人不仅是养老的客体，又是养老的主体，在老年人养老过程中，老年人自身、家庭、社区以及国家都在其中发挥着重要的作用。“中国老龄事业发展‘十二五’规划”中提出的“老有所养、老有所医、老有所教、老有所学、老有所为、老有所乐”的工作目标涵盖了养老的各个方面，但由于河北省农村经济不太发达，农村养老水平较低，所以问卷设计主要侧重于“老有所养”、“老有所医”和老年人精神文明建设这三个方面。在养老主体方面，从老年人自身、老年人家庭、社区和国家四个层面设计问题。

因此问卷主要包括：①老年人个人养老满意度影响因素，包括性别、年龄、文化水平、职业、经济来源、收入、身体健康状况；②老年人家庭建设，包括婚姻状况、子女个数、子女主要供养方式、居住状况；③社区建设状况，包括社区医疗卫生状况和文化娱乐活动状况和满意度；④社会保障状况，包括村集体养老保障建设状况和农村合作医疗实施状况和满意度。具体见表 16-1。

表 16-1　老年人养老满意度问卷主要指标

个人因素	性别 年龄 文化程度 婚姻状况 健康状况 经济来源 职业
家庭因素	子女个数 男孩个数 居住状况 子女照顾方式

续表

社区因素	社区医疗机构方便程度 社区医疗机构医疗水平 社区医疗机构服务水平 社区医疗机构收费水平 社区娱乐活动场所 社会活动
社会保障因素	村集体补贴 医疗保险或新农合落实情况 医疗保险或新农合起付标准 医疗保险或新农合最高自负额 医疗保险或新农合报账比例 医疗保险或新农合报账及时性 低保制度落实情况

由于我国农村普遍是家庭养老或居家养老，所以在设计老年人养老满意度调查问卷的同时，还设计了对应的问卷，主要是针对农村年轻人对老年人供养意愿和能力，以及年轻人对自己养老预期的调查，以形成对应分析。内容包括：年轻人个人基本情况，经济供养来源，居住地选择，精神慰藉需求和个人养老预期。

5．调查结果

调查采用纸质调查和电话调查相结合的方式。电话调查收集老年人问卷和年轻人问卷共 1007 份，其中老年人问卷 108 份，面访调查收集老年人问卷 1716 份，所以本次调查共收集到农村老年人调查问卷 1824 份。

第 17 章　养老满意度数据的预处理及描述统计分析

17.1　数据缺失值的处理

缺失值处理是抽样调查数据处理的一道重要程序。

首先，对数据进行审核，对于缺失值过多的变量(问项)和无效问卷(主要为所有选项均填一项、缺失值过多或者前后逻辑不符等情况)进行删除处理。本调查将回答问项少于问卷总体问项 50%的问卷视为无效问卷，作删除处理，此类问卷共 156 份，另外剔除无效观测 99 份(回答选项均选一项、回答逻辑前后明显不符等情况)，最终得到有效问卷 1569 份，有效问卷回收率 86.02%。问卷回收率较高。

第二，对缺失值进行插补。常见的缺失值插补有回归、列表、均值及算法插补等。由于问卷数据多是分类数据，本调查采用 C&RT 算法填充缺失值。该填充法的一大优点是先根据其他指标对个体进行分类，然后按照分类结果填充缺失值，填充值与原值一致性较高。缺失值处理部分使用的软件是 Clementine 12.0。填充过程如图 17-1 所示。

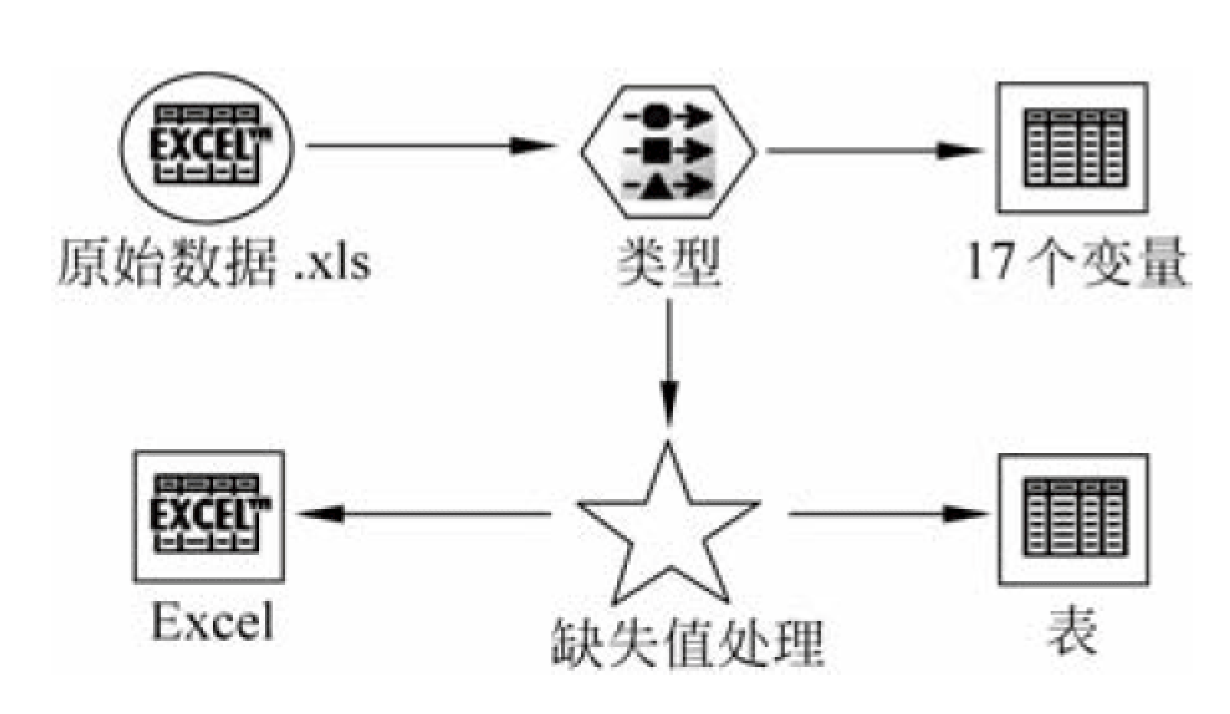


图 17-1　缺失值处理过程图

17.2　调查数据频数分析

1. 个人因素频数分析

从预处理以后的数据频数分布来看，男性人数为 811 人，占调查总人数的 50.8%，女性为 785 人，占调查总人数的 49.2%，调查数据的男女比例合理。

从年龄构成上，50～60 岁人数为 546 人，60～70 岁人数为 425 人，70 岁以上人数为 625 人，各年龄段分布比较均匀。

文化程度方面，没上过学的人数为 482 人，小学文化水平的有 568 人，初中文化水平的有 369 人，高中文化程度的有 157 人，大学及以上文化程度有 20 人。从文化程度的比例来看，河北省第六次人口普查公布的文盲率为 2.61%，而所调查的河北省农村老年人中，没上过学的占总人数的 30.2%，反映出河北省文盲人口中农村老年人占很大的比重；小学和初中文化水平的人数达到总人数的 58.7%，比重最高，可见目前河北省农村老年人口的文化水平普遍偏低。

从职业分布看，务农的人数占到了总调查人数的 59.9%，个体户的比例第二，为 13.6%，外出务工人数为 8.3%，其他的职业比例都很低，这也符合农村的现状。

从经济来源看，经济来源为“个人劳动所得”的人数占总调查人数的 50.8%，“靠子女供养”的为 25.8%，二者之和为 76.6%，这说明目前农村老年人主要是靠自己和子女供养，国家和集体在老年人经济供养方面所起的作用有限。

2. 家庭因素频数分析

从子女个数来看，子女个数在 1～3 个之间的老年人数占调查总数的 89.2%，无子女的老年人只占总体的 9.8%。

从居住状况来看，选择在养老机构居住或其他的人数占总体的 4.3%，比例很低，95.7%的受访老年人都是独自居住或和子女同住的。

3. 社区因素频数分析

在娱乐活动和社会活动方面，分别有29.8%和40%的受访农村老年人有娱乐活动和社会活动，比例偏低。

在村集体补贴和参加农村合作医疗的情况方面，有45.1%的老年人有村集体补贴，受访老年人中参加农村合作医疗比重较高，达到了79.5%。

4. 国家因素频数分析

低保制度的满意度方面，认为“很好”和“好”的比例之和为26.3%，认为“一般”的比例为54.8%，认为“不好”和“很不好”的比例为18.9%，从总体来看，持肯定态度的受访农村老年人比重较高。

5. 总体满意度频数分析

此次受访的农村老年人对自己养老的总体满意度，非常满意的占7.2%，满意的为41.0%，一般的为42.1%，不满意的为8.3%，很不满意的为1.3%，总的来看，农村老年人对养老的满意水平较高。调查数据频数分析见表17-1。

表 17-1　调查数据频数分析

变量名	问卷内容	对应选项	频　数	频率/%
Q_3	性别	男	811	50.8
		女	785	49.2
Q_4	年龄	50～60岁	546	34.2
		60～70岁	425	26.6
		70岁以上	625	39.2
Q_5	文化程度	没上过学	482	30.2
		小学	568	35.6
		初中	369	23.1
		高中	157	9.8
		大学及以上	20	1.3

续表

变量名	问卷内容	对应选项	频数	频率/%
Q_6	职业	务农	956	59.9
		个体户	217	13.6
		外出务工	132	8.3
		公司职员	75	4.7
		公务员	37	2.3
		事业单位	60	3.8
		其他	119	7.5
Q_7	经济来源	个人劳动所得	811	50.8
		养老保险	124	7.8
		离退休金	141	8.8
		政府补贴	32	2.0
		村集体补贴	19	1.2
		子女供养	411	25.8
		其他	58	3.6
Q_8	月平均收入	100 元以下	288	18.0
		101～300 元	387	24.2
		301～500 元	265	16.6
		501～1000 元	220	13.8
		1001～2000 元	232	14.5
		2000 元以上	204	12.8
Q_9	健康状况	非常健康	126	7.9
		健康	690	43.2
		一般	590	37.0
		体弱多病	175	11.0
		生活不能自理	15	0.9
Q_{10}	婚姻状况	未婚	14	0.9
		已婚	1208	75.7
		离异	79	4.9
		丧偶	282	17.7
		其他	13	0.8

续表

变量名	问卷内容	对应选项	频数	频率/%
Q_{11}	子女个数	0 1 2 3 3 个以上	157 484 453 486 16	9.8 30.3 28.4 30.5 1.0
Q_{14}	居住状况	独居或和配偶居住 和子女居住 养老机构居住 其他	770 757 47 22	48.2 47.5 2.9 1.4
Q_{20}	医疗状况	非常满意 满意 一般 不满意 很不满意	11 919 630 36 0	0.7 57.6 39.5 2.3 0.0
Q_{22}	娱乐活动	有 没有	476 1120	29.8 70.2
Q_{23}	社会活动	有 没有	638 958	40.0 60.0
Q_{28}	村集体补贴	有 没有	720 876	45.1 54.9
Q_{29}	农村合作医疗	有 没有	1269 327	79.5 20.5
Q_{31}	低保制度	很好 好 一般 不好 很不好	57 362 874 256 47	3.6 22.7 54.8 16.0 2.9
Q_{33}	总体满意度	非常满意 满意 一般 不满意 很不满意	115 655 672 133 21	7.2 41.0 42.1 8.3 1.3

17.3　变量类别值的调整

从频数分析，我们发现部分变量取值的频数很低，甚至有的变量取值（如 Q_{20}“非常不满意”的选项）的频数为 0，因此应当对这些变量的类别值进行调整。这里用 Clementine 12.0 对变量类别值进行调整。过程和结果如图 17-2 和表 17-2 所示。

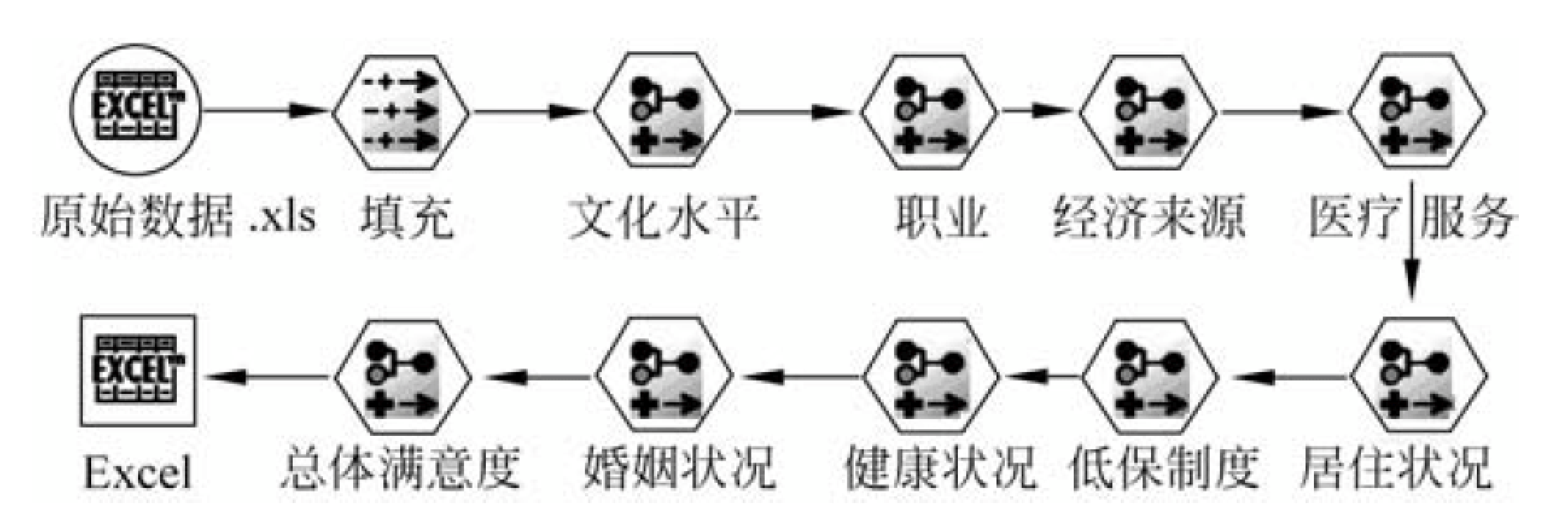


图 17-2　变量类别调整数据流

表 17-2　变量类别值调整表

变量	调整前变量选项及频数	调 整 方 式	调整后变量选项及频数
文化水平	1. 没上过学(482) 2. 小学(568) 3. 初中(369) 4. 高中(157) 5. 大学及以上(20)	1→1 2→2 3、4、5→3	1. 没上过学(482) 2. 小学(568) 3. 初中及以上 (546)
职业	1. 务农(956) 2. 个体户(217) 3. 外出务工(132) 4. 公司职员(75) 5. 公务员(37) 6. 事业单位(60) 7. 其他(119)	1→1 2、3、4、5、6、7→2	1. 务农(956) 2. 非务农(640)

续表

变量	调整前变量选项及频数	调 整 方 式	调整后变量选项及频数
经济来源	1. 个人劳动所得(811) 2. 养老保险(124) 3. 离退休金(141) 4. 政府补贴(32) 5. 村集体补贴(19) 6. 子女供养(411) 7. 其他(58)	1→1 6→2 2、3、4、5、7→3	1. 个人劳动所得(811) 2. 子女供养(411) 3. 保险、补贴及其他(374)
健康状况	1. 非常健康(126) 2. 健康(690) 3. 一般(590) 4. 体弱多病(175) 5. 生活不能自理(15)	1、2→1 3→2 4、5→3	1. 健康(816) 2. 一般(590) 3. 不健康(190)
医疗状况	1. 非常满意(11) 2. 满意(919) 3. 一般(630) 4. 不满意(36) 5. 很不满意(0)	1、2→1 3→2 4、5→3	1. 满意(930) 2. 一般(630) 3. 不满意(36)
居住状况	1. 独居或和配偶居住(770) 2. 和子女居住(757) 3. 养老机构居住(47) 4. 其他(22)	1→1 2→2 3、4→3	1. 独居或和配偶居住(770) 2. 和子女居住(757) 3. 养老机构居住及其他(69)
低保制度	1. 很好 (57) 2. 好(362) 3. 一般(874) 4. 不好(256) 5. 很不好(47)	1、2→1 3→2 4、5→3	1. 好(419) 2. 一般(874) 3. 不好(303)
婚姻状况	1. 未婚 (14) 2. 已婚 (1208) 3. 离异 (79) 4. 丧偶 (282) 5. 其他(13)	2→1 1、3、4、5→2	1. 已婚(1208) 2. 其他(388)

续表

变量	调整前变量选项及频数	调 整 方 式	调整后变量选项及频数
总体满意度	1. 非常满意(115) 2. 满意 (655) 3. 一般 (672) 4. 不满意 (133) 5. 很不满意(21)	1、2→1 3→2 4、5→3	1. 满意(770) 2. 一般(672) 3. 不满意(154)

第 18 章　农村养老满意度的影响因素分析

18.1　基于决策树的养老满意度影响因素分析

18.1.1　变量的重要性分析

通常数据挖掘的数据量较为庞大，在分类预测问题中，变量重要性分析的目的是找到对输出变量的分类预测有重要贡献的变量，保留它们的同时剔除那些对输出变量没有意义的不重要变量。变量的重要性从两个方面联合考虑：①从变量本身考察，重要的变量应携带信息较多，即方差较大。②从变量与输出变量相关角度考察，重要的输入变量应对输出变量的分类预测有显著意义。根据输入和输出变量的计量类型不同，测度输入变量对输出变量的重要性所采用的方法也不同。本章中输入变量和输出变量均为分类变量，利用似然比卡方检验方法，似然比卡方的数学定义为

$$T = 2\sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij}^{0} \ln \frac{f_{ij}^{0}}{f_{ij}^{e}}$$

其中，f^0 为观测频数，f^e 为期望频数。

Clementine 将自动计算似然比卡方的观察值和对应的 1－概率 p 值。1－概率 p 值越高，输入变量与输出变量的相关性越强，输入变量越重要。

本章用 Clementine 对 16 个输入变量的重要性进行分析，过程如图 18-1 所示。条件设定如下：

（1）变量中缺失值所占比例大于 70%应视为不重要变量；

（2）分类变量中类别值所占比例大于 90%应视为不重要变量；

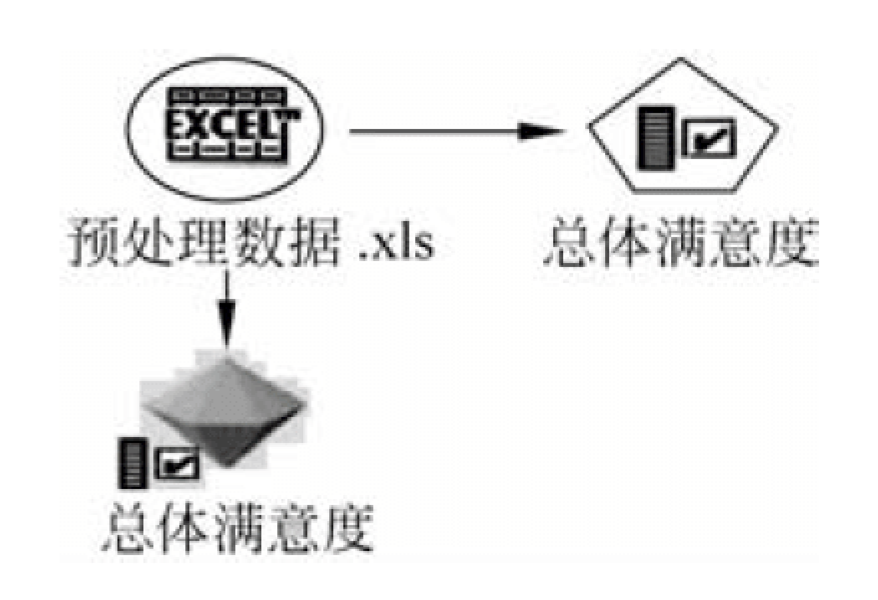


图 18-1　变量重要性分析数据流

(3) 分类变量的类别个数占总样本的比例大于95%应视为不重要变量；

(4) 数值型变量的变异系数小于0.1应视为不重要变量；

(5) 数值型变量的标准差小于0应视为不重要变量。

当输出和输入变量均为分类变量时，按照似然比检验的1—概率 p 值降序排列，在本章中指定1—概率 p 值大于75%为重要变量，结果如表18-1所示。

表 18-1　变量重要性检验

变　　量	1—概率 p 值	变量重要性
低保制度	1.0	重要
农村合作医疗	1.0	重要
居住状况	1.0	重要
娱乐活动	1.0	重要
职业	1.0	重要
月平均收入	0.994	重要
子女个数	0.986	重要
村集体补贴	0.979	重要
经济来源	0.965	重要
文化水平	0.939	重要
社会活动	0.875	重要
年龄	0.769	重要
健康状况	0.559	不重要
婚姻状况	0.502	不重要
性别	0.319	不重要
医疗服务	0.241	不重要

可以看出，前两个最重要的变量为低保制度和农村合作医疗，这两个变量涉及农村社会保障和医疗保障这两个方面，可见社会保障和医疗保障对预测农村老年人养老满意度的重要性最高，这也反映出社会的现实情况：农村经济发展落后，老年人对社会基本保障的需求最为迫切。居住状况对农村老年人养老满意度的重要性排在第三位，可见农村老年人是否与子女同住对农村老年人养老满意度的影响较大。健康状况、婚姻状况、性别和医疗服务的 1－概率 p 值较小，这几个变量对于农村老年人养老满意度的重要性较小。因此为了简化分析，在下文的分析中，可以将对预测农村老年人养老满意度重要性较小的健康状况、婚姻状况、性别和医疗服务这 4 个变量剔除掉。

18.1.2　分类回归树建模分析

1. 问题理解

在对农村养老满意度问卷数据建立决策树模型之前，需要对问题有一个充分的理解。本章对农村养老满意度的决策树分析的目的是，通过决策树挖掘发掘农村老年人中不同养老满意度人群的特征。由于农村养老满意度调查问卷分析的决策树模型比较庞大，因此我们着重解释对农村养老满意度高和不满意度高的人群特点。

本章河北省农村老年人养老满意度人群细分借鉴了市场调查研究中著名的市场细分理论。市场细分的概念是美国市场学家温德尔·史密斯于 20 世纪 50 年代中期提出的，指按照消费者欲望与需求把一个总体市场划分成若干个具有共同特征的子市场的过程。分属于同一细分市场的消费者，他们的需要和欲望极为相似，分属于不同细分市场的消费者对同一产品的需要和欲望存在着明显的差别。

在查阅特定问题下人群细分的诸多研究中，发现常用的细分变量主要是性别、年龄、学历、收入水平、职业、个人所在地等人口统计学细分变量。但随着社会经济的发展，思想文化水平的提高，人们看待问题的观点和侧重点也越来越复杂，仅仅利用这几类变量对农村养老满意度进行人群细分很难看出问题的本质。因此，本章细分变量的确定借鉴了《CHAID 分析在抽油烟机产品市场细分中的应用》(苏胜强，

2006)的思想,结合上文变量重要性的分析,排除了对输出变量农村老年人养老满意度重要性较小的性别、婚姻状况、健康状况和医疗服务,将细分变量分为广度细分变量和深度细分变量,变量分为:广度细分变量为调查对象的基本情况及家庭建设情况,包括年龄、文化程度、职业、经济来源和月平均收入;深度细分变量为反映农村养老状况的家庭情况、精神文化活动、医疗保障和社会保障建设的相关变量,包括子女个数、居住情况、娱乐活动、社会活动、村集体补贴、农村合作医疗、低保制度。

2. 定义数据源

经过预处理之后,得到的数据集为预处理数据.xls,该数据集中包含 1596 条记录,本章选取 13 个变量进行决策树分析,其中输出变量和输入变量如表 18-2、表 18-3、表 18-4 所示。

表 18-2　输出变量

变量名	问卷内容	对 应 选 项
Q_{33}	总体满意度	1. 满意;2. 一般;3. 不满意

表 18-3　输入变量——广度细分变量

变量名	问卷内容	对 应 选 项
Q_4	年龄	1. 50～60 岁;2. 60～70 岁;3. 70 岁以上
Q_5	文化程度	1. 没上过学;2. 小学;3. 初中及以上
Q_6	职业	1. 务农;2. 非务农
Q_7	经济来源	1. 个人劳动所得;2. 子女供养;3. 保险、补贴及其他
Q_8	月平均收入	

表 18-4　输入变量——深度细分变量

变量名	问 卷 内 容	对 应 选 项
Q_{11}	子女个数	1. 0 个;2. 1 个;3. 2 个;4. 3 个;5. 3 个以上
Q_{14}	居住状况	1. 独居或和配偶居住;2. 和子女居住;3. 养老机构居住及其他

续表

变量名	问卷内容	对应选项
Q_{22}	娱乐活动	1. 有；2. 没有
Q_{23}	社会活动	1. 有；2. 没有
Q_{28}	村集体补贴	1. 有；2. 没有
Q_{29}	农村合作医疗	1. 有；2. 没有
Q_{31}	低保制度	1. 好；2. 一般；3. 不好

3. 建立模型

用 Clementine 的分类回归树对数据进行分类研究。

第一，读入预处理数据.xls 并定义数据类型。

第二，过滤变量，加入 Filter 节点，将对输出变量农村老年人养老总体满意度重要性较小的健康状况、婚姻状况、性别和医疗服务这四个变量剔除掉。

第三，样本分割，将样本分割成训练集和测试集，其中训练集用来建立模型，验证集用来对所建模型进行评估。加入 Sample 节点作随机抽样，固定随机数种子(7 956 454)，由于样本量不是很大，为了使模型更趋于准确，将数据分为训练集和验证集两部分，其中训练集为 85%，验证集为 15%。

第四，加入 Filter 节点，过滤掉深度细分变量，建立广度细分模型。

第五，加入 Filter 节点，过滤掉广度细分变量，建立深度细分模型。

第六，对模型进行评价。

具体的数据流如图 18-2 所示。

4. 模型设定

在利用 CART 算法建立决策树的过程中，首先使用 CART 简单模式对数据进行探索性分类(具体结果不再赘述)，第二步使用专家模式，具体设定如下：树的最大深度设为 5(即根节点下的最大层数为 5)，最大代用项为 5；设定分类目标的杂质项测量系数为 Gini 系数，并指定在树的生长过程中进行新的分割所需的最小杂质改变量为

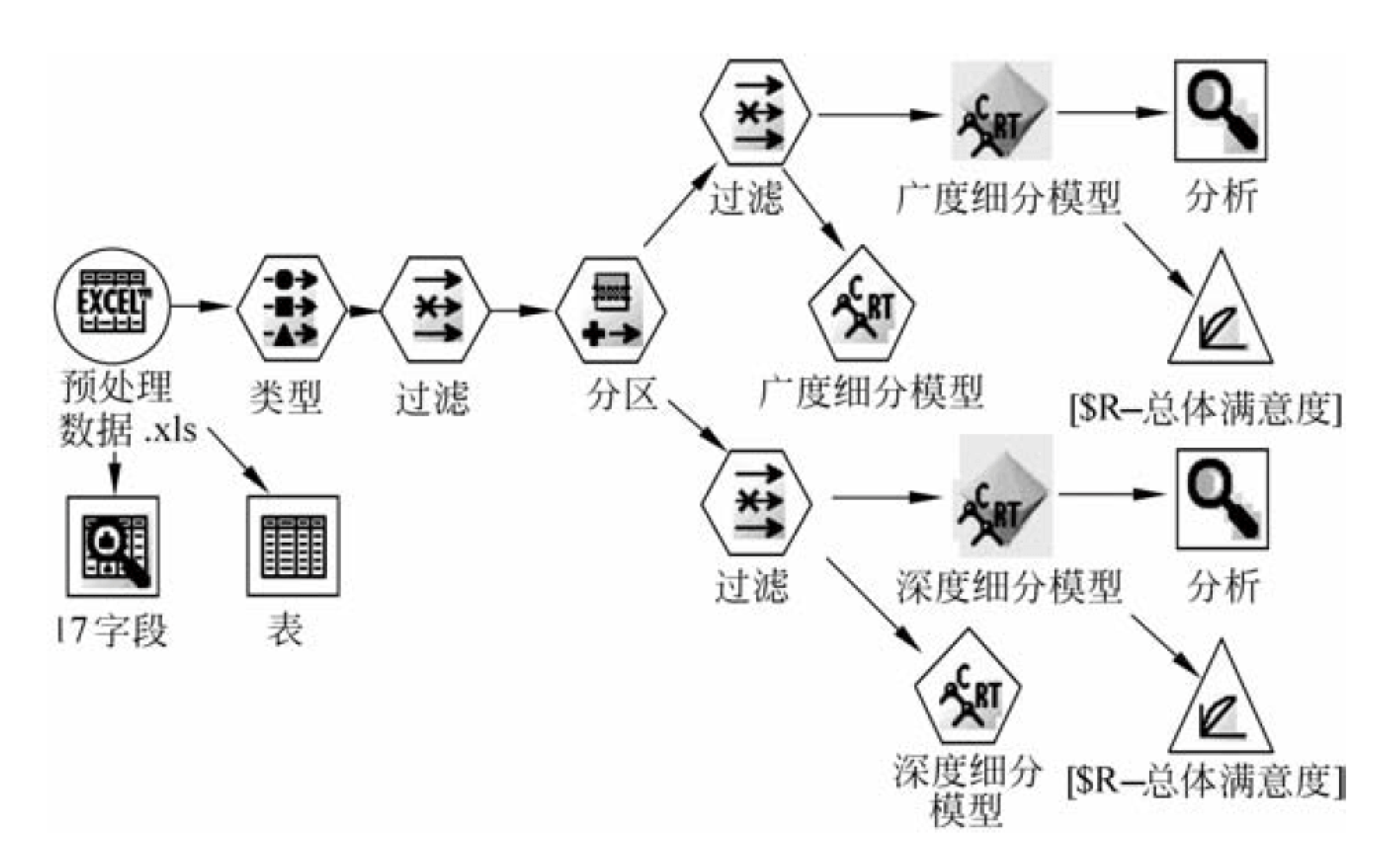


图 18-2　决策树分析数据流

0.0001，若某一分支的最佳分割所引起的杂质改变量低于 0.0003，则决策算法程序不会继续进行分割；选择修建树选项，即当进行一个节点的分支的构建时，如不能显著提高树的准确性，则自动剔除此节点以下的所有分支；使用百分比的方法来控制树的生长，设定父节点记录数占总记录数为最少 2%，子节点记录数占总记录数为最少 1%。在此设定下，得到最终决策树结果。

5. 模型结果

执行数据流，得到广度细分分类回归树模型和深度细分分类回归树模型的输出结果。

表 18-5 显示的是广度细分变量的重要性大小依次为职业、月平均收入、文化水平、经济来源和年龄，其中年龄对农村老年人养老满意度的重要性很小，仅为 0.14。

表 18-5　广度细分变量重要性

变　　量	重要性	变　　量	重要性
职业	0.27	经济来源	0.19
月平均收入	0.21	年龄	0.14
文化水平	0.2		

图 18-3 为广度细分 CART 决策树，这是一棵五层决策树，根节点包含 1596 个样本，养老满意度满意的人数为 770 人，一般的为 672 人，不满意的为 154 人，相应的百分比为 48.24%、42.11%和 9.65%。决策树的第一最佳分组变量为职业，并以此形成二叉树。由于决策树层次较多，分类比较复杂，本章不再一一分析。根据决策树及预测规则，最终得出的分析结论如下：

职业为务农，月平均收入在 501～1000 元之间，则养老满意度为满意，样本量为 138 个，置信度为 0.63；

职业为务农，月平均收入在 500 元以下或 1000 元以上，经济来源为子女供养，文化水平为初中及以上，则养老满意度为一般，样本量为 29 个，置信度为 0.517；

职业为务农，月平均收入在 500 元以下或 1000 元以上，经济来源为子女供养，文化水平为小学及以下，年龄在 50～60 岁之间，则养老满意度为一般，样本量为 29 个，置信度为 0.517；

职业为务农，月平均收入在 500 元以下或 1000 元以上，经济来源为子女供养，文化水平为小学及以下，年龄在 60 岁以上，则养老满意度为满意，样本量为 243 个，置信度为 0.572；

职业为务农，月平均收入在 500 元以下或 1000 元以上，经济来源为个人劳动所得或保险、补贴及其他，年龄在 60～70 岁之间，则养老满意度为一般，样本量为 105 个，置信度为 0.457；

职业为务农，月平均收入在 500 元以下或 1000 元以上，经济来源为个人劳动所得或保险、补贴及其他，年龄在 50～60 岁之间或 79 岁以上，则养老满意度为满意，样本量为 412 个，置信度为 0.507；

职业为非务农，月平均收入在 100 元以下，则养老满意度为满意，样本量为 45 个，置信度为 0.644；

职业为非务农，月平均收入在 100～300 元之间或 1000 元及以上，则养老满意度为一般，样本量为 434 个，置信度为 0.518；

职业为非务农，月平均收入在 300～1000 元之间，经济来源为个人劳动所得，则养老满意度为满意，样本量为 73 个，置信度为 0.479；

职业为非务农，月平均收入在 300～1000 元之间，经济来源为子女

供养或保险、补贴及其他，文化水平为小学以下，则养老满意度为满意，样本量为 17 个，置信度为 0.529；

职业为非务农，月平均收入在 300～1000 元之间，经济来源为子女供养或保险、补贴及其他，文化水平为小学及以上，则养老满意度为一般，样本量为 71 个，置信度为 0.521。

如表 18-6 所示，深度细分变量对养老满意度的重要性大小依次为低保制度、农村合作医疗、村集体补贴、社会活动、居住状况、娱乐活动和子女个数，其中子女个数对农村老年人养老满意度的重要性很小。

表 18-6 深度细分变量重要性

变　　量	重要性	变　　量	重要性
低保制度	0.8	居住状况	0.02
农村合作医疗	0.13	娱乐活动	0.01
村集体补贴	0.03	子女个数	0
社会活动	0.02		

图 18-4 为深度细分决策树，一共有五层，根节点与广度细分决策树相同，不再赘述。深度细分决策树的第一最佳分组变量为低保制度，以此形成二叉树。根据决策树及预测规则，最终得出的分析结论如下：

低保制度实施情况的评价为“好”，则养老满意度为满意，样本量为 419 个，置信度为 0.745；

低保制度实施情况的评价为“一般”，参加农村合作医疗，子女个数为 0 个、1 个或 3 个，则养老满意度为满意，样本量为 486 个，置信度为 0.502；

低保制度实施情况的评价为“一般”，参加农村合作医疗，子女个数为 2 个或 3 个以上，则养老满意度为一般，样本量为 201 个，置信度为 0.507；

低保制度实施情况的评价为“一般”，没有参加农村合作医疗，居住状况为独居或和配偶居住，或者是养老机构居住及其他，则养老满意度为一般，样本量为 90 个，置信度为 0.678；

低保制度实施情况的评价为“一般”，没有参加农村合作医疗，居住状况为和子女同住，一年内没有参加社会活动，则养老满意度为满意，样本量为 67 个，置信度为 0.478；

低保制度实施情况的评价为“一般”，没有参加农村合作医疗，居住状况为和子女同住，一年内参加社会活动，则养老满意度为一般，样本量为 30 个，置信度为 0.633；

低保制度实施情况的评价为“不好”，参加农村合作医疗，有村集体补贴，则养老满意度为一般，样本量为 85 个，置信度为 0.565；

低保制度实施情况的评价为“不好”，参加农村合作医疗，没有村集体补贴，子女个数为 3 个，则养老满意度为满意，样本量为 38 个，置信度为 0.474；

低保制度实施情况的评价为“不好”，参加农村合作医疗，没有村集体补贴，子女个数为 2 个及以下，则养老满意度为一般，样本量为 98 个，置信度为 0.48；

低保制度实施情况的评价为“不好”，没有参加农村合作医疗，居住状况为养老机构居住或其他，则养老满意度为不满意，样本量为 21 个，置信度为 0.762；

低保制度实施情况的评价为“不好”，没有参加农村合作医疗，居住状况为独居或和配偶居住，或者是与子女同住，子女个数为 1 个，则养老满意度为不满意，样本量为 18 个，置信度为 0.444；

低保制度实施情况的评价为“不好”，没有参加农村合作医疗，居住状况为独居或和配偶居住，或者是与子女同住，子女个数为 0 个或 2 个及以上，则养老满意度为一般，样本量为 43 个，置信度为 0.535。

6. 模型评估

模型评估是数据挖掘过程中的关键一步，只有通过了评估才能确定前面的各步骤已经成功地完成。因而，在模型建立之后，要对模型进行评估，以得到更合理、更完备的决策信息。下面主要采用建立模型收益图和计算数据预测结果进行对比来进行模型评估。

收益图的对角线表示的是如果不使用模型，整个样本的期望响应，而收益曲线越陡，增益就越高。如果不使用模型，调查人数与所获得的

收益同比例增长，即在调查对象中，选项为“满意”的人数占总人数的比例与所调查的人数比例相等。图 18-5 和图 18-6 分别为广度细分模型和深度细分模型的收益图，曲线表示，随着样本量的增加，基于增益能多大程度地提高响应率。以深度细分模型中总体满意度为“比较满意”为例，由模型收益图可以看出，本章设定的模型对于修正总体具有一定的提升意义。

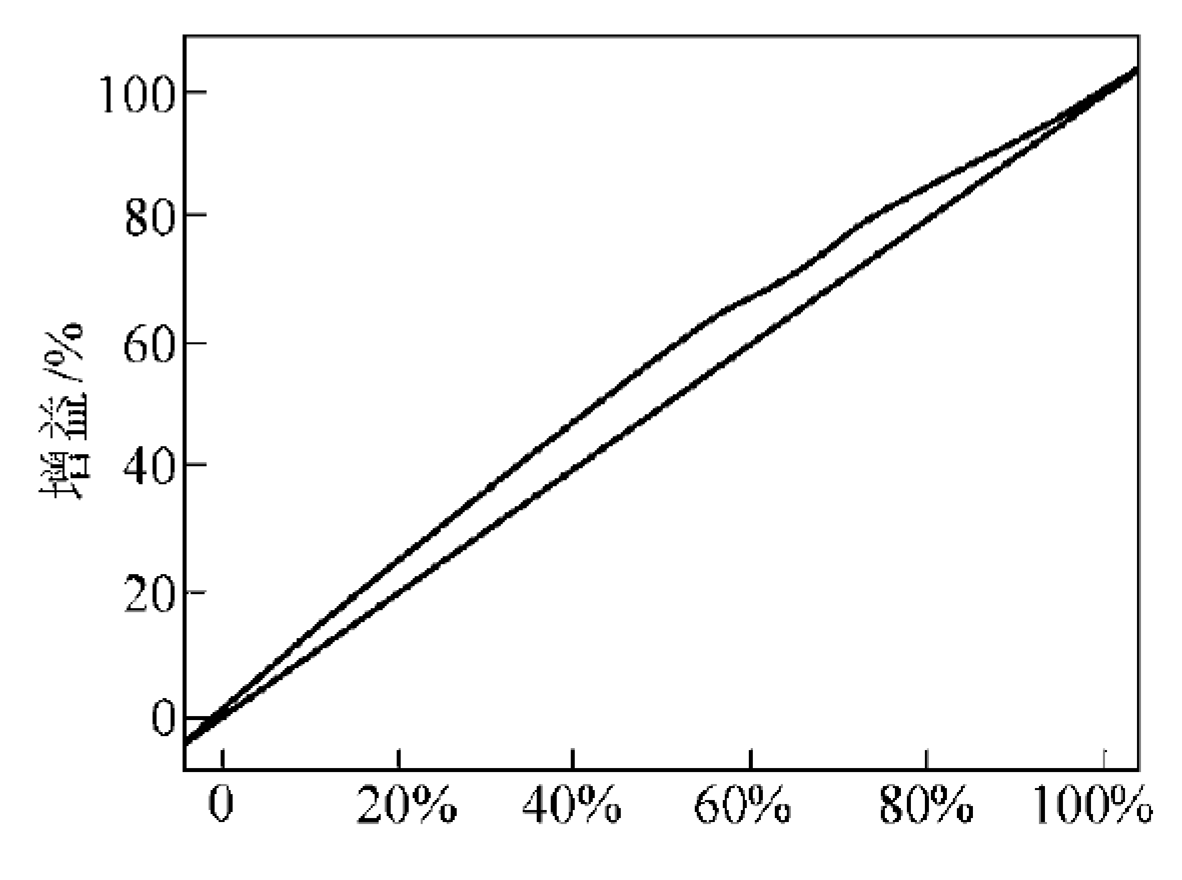


图 18-5　广度细分模型累计收益图

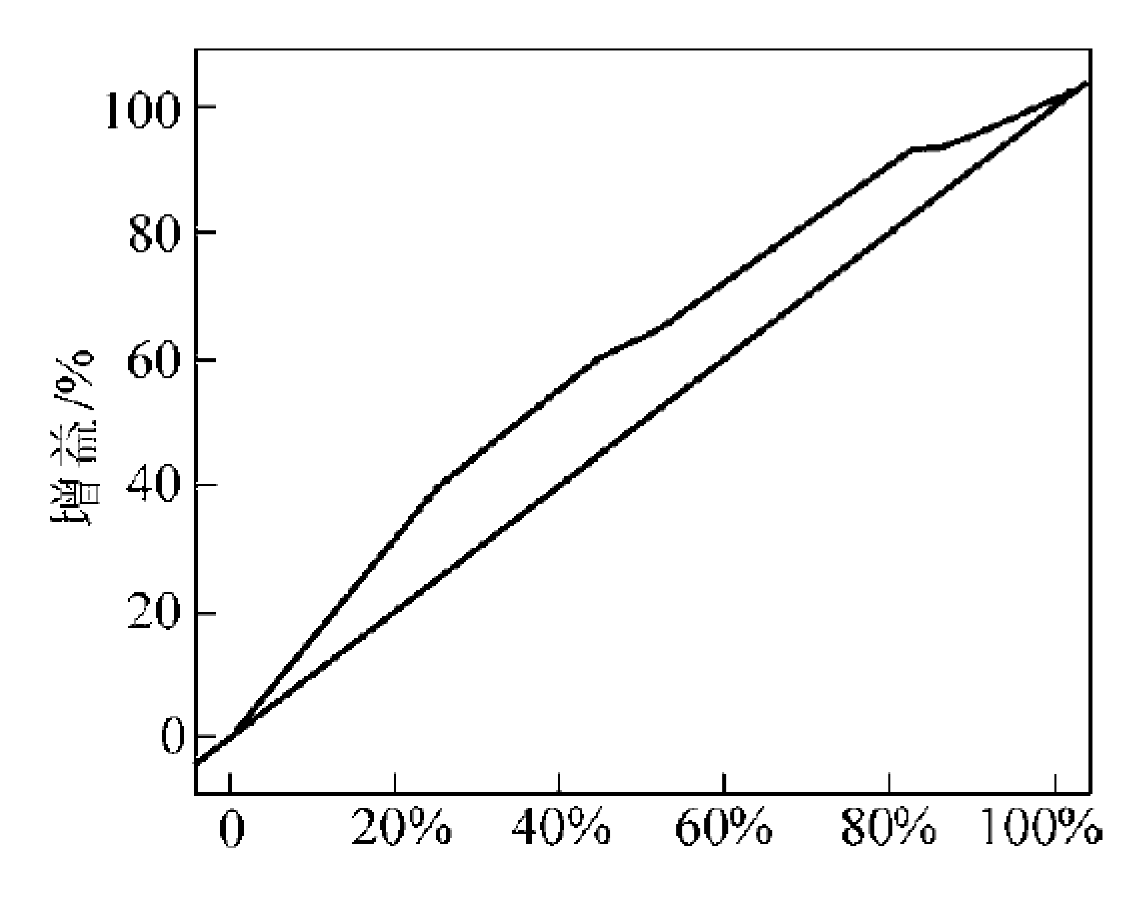


图 18-6　深度细分模型累计收益图

表 18-7 为广度细分模型和深度细分模型的预测结果，两个模型的预测准确率都不太高，但达到了一定的比率。主要原因是调查数据量不足。

表 18-7 模型预测结果

类别	广度细分预测结果		深度细分预测结果	
	样本量	比例/%	样本量	比例/%
正确	848	53.13	930	58.27
错误	748	46.87	666	41.73

7. 模型结果分析

综上可以看出，影响农村老年人养老满意度的个人因素中，性别、婚姻状况和健康状况对其影响较小，农村老年人养老满意度在这几个因素上的差异性很小，职业、月平均收入、文化水平、经济来源和年龄这五个对养老满意度的重要性依次递减。调整后的调查数据中，职业分为务农和非务农，职业反映了一个人的社会经历，个人的经历对老年人养老满意度的影响较大；月平均收入对养老满意度的影响规律并不是很明显，因为作为反映经济状况的变量，个人储蓄更具代表性，月平均收入并不能完全反映出老年人的经济状况，但由于个人储蓄是一个敏感话题，本次调查中没能准确收集到这一变量的数据；在经济来源方面，大体可以看出经济来源独立的老年人满意度较高；年龄对老年人养老满意度的重要性较小，可以看出，年龄越大老年人的满意度越高，由于满意度是一个心理因素，这种情况的出现可能是与年老的老年人的养老需求较低有关。

在影响农村老年人养老满意度的家庭因素中，子女个数对养老满意度的重要性不很高，而居住状况对农村老年人的养老满意度有很大的影响，在决策树分析中，我们看到，居住状况为养老机构居住或其他的 21 个样本中，有 16 个样本对养老满意度的评价为不满意，比重达到了 76.19%。

在影响农村老年人养老满意度的社区因素中，社区医疗机构医疗水平对农村老年人的养老满意度影响很小，这是由于我们广大农村地区医疗水平普遍偏低，差异性较小，区分度不明显。因此我们不能得出结论，认为社区医疗水平对提高农村老年人养老满意度不重要。社区娱乐活动和老年人社会活动这两个因素对老年人养老满意度有一定的

影响，但是决策树分析中反映出的不完全是正向的影响，这说明社区建设中，这两方面存在着不足。

在影响农村老年人养老满意度的社会保障和医疗保障因素中，低保制度的实施情况的好坏对农村老年人养老满意度高低的影响很大，这是因为，低保制度的实施状况从侧面反映了当地社会保障水平的高低，社保水平高则老年人的养老满意度高，社保水平低则相反；是否参加农村合作医疗是仅次于低保制度实施状况的重要变量，参加农村合作医疗的老年人养老满意度较高。

总之，影响老年人养老满意度的个人因素、家庭因素、社区因素以及社会保障和医疗保障因素是相互影响、共同起作用的。

18.2　农村养老满意度的关联分析

本部分分析的变量中包含非布尔类型数据，因此采用 GRI 算法。

1. 变量前后项设定

经过预处理后的数据由 Clementine 数据挖掘软件中的 GRI 模型来实现挖掘过程。在进行挖掘分析前，须定义变量类型。如表 18-8 所示，职业、经济来源、居住状况、娱乐活动、社会活动、村集体补贴和农村合作医疗等变量均为有限个值，且各值之间无序，因此定义成标称类型；年龄、月平均收入、文化水平和总体满意度各变量值之间具有一个隐含的序，因此定义成连续型数值类型；同时指定总体满意度为关联规则的后件，其余为关联规则的前件。

表 18-8　关联规则变量

变　　量	调整前变量选项	类型	前项/后项
职业	1. 务农；2. 非务农	标称	前项
月平均收入	1. 100 元以下；2. 101～300 元；3. 301～500 元；4. 501～1000 元；5. 1001～2000 元；6. 2000 元以上	数值	前项
文化水平	1. 没上过学；2. 小学；3. 初中及以上	数值	前项

续表

变　　量	调整前变量选项	类型	前项/后项
经济来源	1. 个人劳动所得；2. 子女供养；3. 保险、补贴及其他	标称	前项
年龄	1. 50～60 岁；2. 60～70 岁；3. 70 岁以上	数值	前项
居住状况	1. 独居或和配偶居住；2. 和子女居住；3. 养老机构居住及其他	标称	前项
娱乐活动	1. 有；2. 没有	标称	前项
社会活动	1. 有；2. 没有	标称	前项
村集体补贴	1. 有；2. 没有	标称	前项
农村合作医疗	1. 有；2. 没有	标称	前项
总体满意度	1. 满意；2. 一般；3. 不满意	数值	后项

2. 关联规则挖掘分析

经过预处理后的数据由 Clementine 12.0 数据挖掘软件中的 Apriori 模型来实现挖掘过程。在关联规则挖掘中需要设置相关阈值。其中，最小支持度和置信度阈值的大小直接影响着关联规则挖掘出来的规则的数量。如果对最小支持度和置信度阈值不加限制，产生的规则数量将会十分惊人，结果将毫无意义。相关文献中并没有明确的依据确定如何设置模型的最小支持度阈值和置信度阈值，本章在实际操作中，采用设置不同最小支持度阈值的方式进行反复挖掘，即进行若干次探索性的挖掘。由结果可知，如果最小支持度阈值设得太低，有可能产生大量冗余规则；如果最小支持度阈值过高，则很可能损失有价值的关联规则信息，最小支持度和置信度阈值的调整明显影响着关联规则的数量。由于题项众多，为了避免过于繁多的规则被发现，经过多次比较发现当最小支出度阈值小于 5%，后件为{总体满意度＝3}的所有模式都将损失掉，因此我们最终选择分析在最小支持度阈值为 5%、最小置信度阈值为 50%的情况下进行关联规则挖掘。

下面对各个后项变量进行分析。本研究从诸多提升度大于 1 的关

联规则中按照最小支持度进行排序，选出排名靠前且有代表性的规则，由于规则众多，在确保支持度大于 5％的条件下，选取部分置信度大于 50％的规则进行论述，选出的规则详见表 18-9、表 18-10、表 18-11。

表 18-9　后件{总体满意度＝1}的关联规则（部分）

规则	前　　项	支持度/％	置信度/％	提升度/％
1	低保制度＝1	26.25	74.46	1.54
2	文化水平＝1 and 居住状况＝2	14.79	61.44	1.27
3	经济来源＝2 and 居住状况＝2	12.28	60.2	1.25
4	村集体补贴＝2 and 低保制度＝1	11.78	74.47	1.54
5	职业＝2 and 低保制度＝1	9.84	70.06	1.45

表 18-10　后件{总体满意度＝2}的关联规则（部分）

规则	前　　项	支持度/％	置信度/％	提升度/％
6	农村合作医疗＝2 and 低保制度＝2	11.72	59.36	1.40
7	职业＝2 and 村集体补贴＝2 and 低保制度＝2	11.72	57.75	1.37
8	文化水平＝3 and 社会活动＝2.0 and 低保制度＝2	11.03	57.39	1.36
9	居住状况＝1 and 娱乐活动＝2 and 社会活动＝2.0 and 低保制度＝2	13.22	55.92	1.33
10	职业＝2 and 娱乐活动＝2 and 社会活动＝2.0 and 低保制度＝2	10.78	55.81	1.33
11	居住状况＝1 and 娱乐活动＝2 and 低保制度＝2	18.36	55.63	1.32
12	文化水平＝2 and 村集体补贴＝2 and 低保制度＝2	11.15	54.49	1.29

表 18-11　后件{总体满意度＝3}的关联规则

规则	前项	支持度/％	置信度/％	提升度/％
13	农村合作医疗＝2 and 低保制度＝3	5.14	52.44	5.4

后项{总体满意度＝1}的关联规则如表 18-9 所示。

规则 1 说明低保制度满意度较高的农村老年人中，养老满意度较高，其概率为 77.46%；

规则 2 说明文化程度较低，和子女同住的农村老年人，养老满意度较高，其概率为 61.44%；

规则 3 说明经济来源为子女供养且和子女同住的农村老年人，养老满意度较高，概率为 60.2%；

规则 4 说明所在地没有村集体补贴，对低保制度实施状况满意的农村老年人养老满意度较高，概率为 74.47%；

规则 5 说明职业为务农，对低保制度实施状况满意的农村老年人养老满意度较高，概率为 70.06%。

综合来看，所在地低保制度实施状况较好，居住状况为和子女同住的农村老年人，养老满意度普遍偏高，另外，文化水平较低的老年人养老满意度较高。

后项{总体满意度=2}的关联规则如表 18-10 所示。

规则 6 说明对没有参加农村合作医疗，对低保制度实施状况的评价为一般的农村老年人养老满意度为一般，其概率为 59.36%；

规则 7 说明职业为非务农，所在地没有村集体补贴，对低保制度实施状况的评价为一般的农村老年人养老满意度为一般，其概率为 57.75%；

规则 8 说明文化程度较高，一年之内没有参加娱乐活动，对低保制度实施状况的评价为一般的农村老年人养老满意度为一般，其概率为 57.39%；

规则 9 说明居住状况为独居或和子女配偶居住，一年之内没有参加娱乐活动和社会活动，且对低保制度实施状况的评价为一般的农村老年人养老满意度为一般，其概率为 55.92%；

规则 10 说明职业为非务农，一年之内没有参加娱乐活动和社会活动，且对低保制度实施状况的评价为一般的农村老年人养老满意度为一般，其概率为 55.81%。

规则 11 说明居住状况为独居或(和)子女配偶居住，一年内没有参加娱乐活动，对低保制度实施状况的评价为一般的农村老年人养老满

意度为一般，其概率为 55.63%。

规则 12 说明文化水平为小学，所在地没有村集体补贴，对低保制度实施状况的评价为一般的农村老年人养老满意度为一般，其概率为 54.49%。

综合来看，所在地低保制度实施状况一般，没有村集体补贴，居住状况为独居或和配偶居住农村老年人养老满意度较低，且文化水平较高，职业为非务农的老年人，养老满意度的水平偏低。另外，我们还可以看出，是否参与社会活动和娱乐活动，也与老年人养老满意度的高低有很大关联。

后项{总体满意度＝3}的关联规则如表 18-11 所示。由于调查数据中养老满意度为不满意的样本数据较少，因而后项{总体满意度＝3}的关联规则的支持度较低，在 5%的支持度阈值下只有一条关联规则。从规则 13 我们可以看到，没有参加农村合作医疗，且对所在地低保制度实施状况评价为不好的农村老年人，养老满意度为不满意，其概率为 52.44%。这体现出社会保障制度和医疗保障制度实施状况的好坏严重影响着农村老年人的养老满意度。

一般来说，所在地低保制度实施状况好，参加农村合作医疗，有村集体补贴，则老年人的养老满意度较高；与子女同住老年人的养老满意度较高；文化水平较高，职业为非务农的老年人养老满意度较低。

第 19 章　农村养老模式的探讨及建议

基于第 18 章对河北省农村老年人养老满意度调查数据的分析,在探讨了影响老年人养老满意度因素的影响程度和相互之间关联关系的基础上,本章将从提高农村老年人养老满意度的角度探讨适合农村老年人的养老模式及相应的可行性建议。

19.1　农村养老模式的探讨

当前我国养老方式主要分为家庭养老、居家养老、社区养老、机构养老和社会养老。根据对已有文献的研究和本次调查数据的分析,传统的家庭养老已经不能满足当今农村老年人的养老需求,农村老年人的养老逐步走出家庭,向社区和社会拓展。基于以上分析和“中国老龄事业发展‘十二五’规划”的工作要求,本章提出适合河北省农村老年人的养老模式为:以居家为基础、社区为依托、机构为支撑,社会养老保障制度和医疗保障制度相结合的养老模式。

1. 居家养老适合河北省农村老年人的养老意愿

从老年人自身因素来看,经济来源反映出老年人对养老经济供养来源的要求,调查数据分析显示,老年人中经济来源为“个人劳动所得”的养老满意度较高,这反映了当前农村老年人养老的经济供养来源更加独立,主要依靠自身。另外,我们还看到老年人养老满意度的高低受子女个数多少的影响很小,但是分析显示,居住方式为“与子女同住”的老年人养老满意度较高。可见,农村老年人养老的经济来源更加独立,对家庭的经济供养依赖性降低,但对在家庭居住的依赖性还很高,因而,在传统的家庭养老模式发展的基础上,居家养老更能提高农村老年

人的养老满意度。

2. 广大农村老年人对社会养老有较高的要求

社会养老主要表现在两个方面：机构养老和社会养老保险制度。本次调查显示老年人倾向于与子女一起居住，居住方式为“养老机构居住”的农村老年人满意度普遍偏低。另一方面，我国农村地区养老保障制度还很不完善，农村老年人的养老满意度的高低很大程度上受到养老保障制度的建设和实施状况的影响。丰富和完善农村社会保障制度，并保证其顺利实施，是普遍提高农村老年人的养老满意度的有效途径。

3. 社区完善的医疗保障制度能够提高农村老年人的养老满意度

从本次调查分析可以看出，参加农村合作医疗的老年人养老满意度较高。目前农村居民参保农村合作医疗制度和农村医疗保险制度的比重在不断提升，但是还不能覆盖全部农村居民，本次调查的样本中有20.5%的农村老年人没有参加农村合作医疗。因而，进一步加大农村合作医疗制度的宣传，尤其是在农村老年人中的宣传，逐步加强和完善农村医疗保障制度的建设是我们国家一项长期而艰巨的任务。

总之，应以居家为基础、社区为依托、机构为支撑，大力加强农村社会养老保障制度建设和完善医疗保障制度，充分调动起农村老年人、家庭、社区和社会的积极性，促进农村老年人家庭建设和老年经济发展，为广大农村老年人创造一个适合的养老环境。

19.2　关于农村养老问题的对策建议

随着我国老龄化社会的不断发展，老龄化问题不断出现新的挑战，为了适应养老需求的变化，缓解中国老龄化社会的压力，为我国广大农村地区老年人营造一个适合的养老环境，实现“老有所养、老有所医、老有所教、老有所学、老有所为、老有所乐”的目标，需要全社会的共同努力。

1. 促进农村经济发展，建设农村老年宜居社区

一方面，我国农村地区经济发展普遍落后，农村年轻人外出打工，农村有大量的留守儿童和留守老人。建设适合老年人居家养老的家庭环境，需要采用经济的手段，政府和社会应大力发展农村经济，吸引农村年轻劳动力回流，为建设适合农村老年人养老的家庭环境提供必要的物质基础。另一方面，在促进农村城镇化建设的同时，建设农村老年宜居社区，增加老年文化、教育和体育健身活动设施，开展无障碍设施改造和老龄设备配套建设，为老年人提供舒适的生活娱乐场所，提高农村老年人的精神文化生活。

2. 加强老年家庭建设，支持居家养老

一要改善老年人居住条件。引导开发老年宜居住宅和代际亲情住宅，鼓励家庭成员与老年人共同生活或就近居住。推动和扶持老年人家庭的无障碍改造。二要完善家庭养老支持政策。完善老年人口户籍迁移管理政策，为老年人随赡养人迁徙提供条件。建立健全家庭养老保障和照料服务扶持政策，完善农村计划生育家庭奖励扶助制度和计划生育家庭特别扶助制度。三要弘扬中华民族孝亲敬老的传统美德。强化尊老敬老道德的建设，提倡亲情互助，营造温馨和谐的家庭氛围，发挥家庭养老的基础作用。建设适合老年养老的温馨家庭，提高老年人居家养老的幸福指数。

3. 促进老龄产业发展，建设多元化养老机构

传统的家庭照料已经越来越不能满足老年人的养老需求，社会提供养老服务越来越为广大群众所认可和接受。为了给老年人提供更好的服务和产品，就要大力促进老龄产业的发展。在全面推进老龄产业发展的进程中，必须统筹兼顾供给与需求两方面，首先要了解老年人的需求，以此制定市场长期发展规划；二是建立一支高素质的从业人员队伍，为老年人提供优质的服务和产品；三是引导老年人转变养老观念和消费观念，将老龄服务和产品带到老年人的身边。另一方面，尽管农村老年人对养老机构居住的认可度较低，但是养老机构是社会养老的重要组成部分，可以为生活不能自理、需要专业照顾的老年人提供完善的

日常照料和医疗护理服务,具有不可替代的作用,因而要农村地区建立多元化的养老机构,满足农村老年人不同层次的养老需求。

4. 完善农村社会保障和医疗保障制度,促进政策实施

建立和健全农村社会保障体系,消除社会保障制度城乡二元化的差异,是国家长治久安和社会稳定的保证。我国正在加大财政改革力度,逐步加大财政对社会保障和医疗保障的转移支付,国家财力要从经济建设支持向公共事业、社会发展和社会保障支持转变,逐步建立和完善公共财政框架。这就为建立农村社会养老保障提供了适合的经济条件和丰富稳定的资金来源,再加上农民有参加社会养老保险的积极意愿,因此在农村尽快建立全面养老保障体系是我国一项首要任务。建立完善的社会保障体系是国家实现再分配的主要方式之一,建立农村社会养老保障体系可以化解许多社会矛盾,促进社会和谐发展。同时只有解决了农民养老的后顾之忧,才会使农民放心地去消费和投资,拉动社会经济的发展。

总之,就是要积极调动个人、家庭、社区和国家的力量,建立一个以居家为基础、社区为依托、机构为支撑,农村社会养老保障制度和医疗保障制度完善的农村养老模式,为我国广大农村地区老年人营造一个适合的养老环境,实现“老有所养、老有所医、老有所教、老有所学、老有所为、老有所乐”的目标。

附录A　新农村建设情况调查

新农村建设情况调查问卷

________县（市、区）________乡（镇）________村

同志，您好！

建设社会主义新农村是构建和谐社会的基础，是统筹城乡发展，全面实现小康社会的重要途径。为了解当前新农村建设的现状，为省委、省政府指导“三农”工作提供决策依据，现进行新农村建设和新民居建设情况调查。请您大力支持、积极配合。谢谢！

填 卷 说 明

一、本问卷的问题都有几个答案，请调查员同调查对象座谈后选择符合调查对象意愿或情况的答案，并在选定的答案前“□”内打“√”，表示回答。除规定选择答案数量的题外，其余一律选择一个答案。

二、名词解释：

“二十字方针”：指中央对我国新农村建设提出“生产发展、生活富裕、乡风文明、村容整洁、管理民主”的建设要求。

“新民居建设”：即农村新民居建设，是改善农村生活环境，提升农民生活质量，增进农民福祉的一件惠民大事，对于统筹城乡协调发展，全面加快社会主义新农村建设具有重要战略意义。只有从实际出发，坚持科学发展，坚持以人为本，才能稳步推进新民居建设进程，推动农村经济社会各项事业科学、协调、全面发展。

调查对象基本情况

一、您的性别为：
□男　□女
二、您的年龄为：
□30 岁以下　□30～45 岁　□45～60 岁
□60 岁以上
三、您的文化程度为：
□不识字或少识字　□小学　□初中
□高中　□大专及以上
四、您的职业是：
□农民群众　□村干部　□乡(镇)干部
□其他
五、您的政治面貌为：
□中共党员　□民主党派　□其他
六、您所在的村是：
□新农村示范村　□新农村达标村　□移民搬迁村
□其他村
七、您现在居住的房子属于以下哪种类型：
□租房子住　□自家的土坯房
□自家砖瓦结构平房　□自建小楼
□政府统一建造的居民区

对农村目前现状的评价

八、您知道什么是社会主义新农村(“二十字方针”)吗?
□知道　□知道一点　□不知道
九、您认为开展新农村建设与自己：
□密切相关　□相关　□不相关
□不清楚
十、您如何看待政府提出的“建设新民居”的政策：
□非常支持　□支持
□先观望一下形势再做表态　□没有关注过,不了解
□无所谓
十一、您对您所在的村(镇)目前农村现状的看法
1. 总体来说,您对当前的各个方面是否满意?
□非常满意　□比较满意　□不太满意

续表

□不满意
2. 您所在村(镇)的养殖业情况为:
□没有人养殖　□散户养殖为主　□规模养殖为主
3. 您所在村的自然条件是:
□居住分散、交通不便、信息闭塞
□居住相对集中、交通、信息还算不错
□居住集中、交通便捷、信息畅通
4. 您对您家近几年的家庭收入提高状况是否满意?
□非常满意　□比较满意　□不太满意
□不满意
5. 您对现在的居住环境是否满意?
□非常满意　□比较满意　□不太满意
□不满意
6. 您认为您所在村(镇)农民的整体文化水平如何?
□水平低　□一般　□水平较高
7. 您认为自己的孩子接受多高的教育就可以了?
□识字就可以了　□初中　□高中
□大学　□大学以上
8. 您对村里的学校的教学条件是否满意?
□非常满意　□比较满意　□不太满意
□不满意
9. 目前您最关心的教育问题是:
□学校的教学环境　□教师素质　□学费
□教育体制改革
10. 总的来说,您对目前农村的教育环境是否满意?
□非常满意　□比较满意　□不太满意
□不满意
11. 您对你所在村(镇)农民医疗卫生、养老、低收入等社会保障是否满意?
□非常满意　□比较满意　□不太满意
□不满意
12. 您对您所在村(镇)农村社会风气是否满意?
□非常满意　□比较满意　□不太满意
□不满意
13. 您对您所在村(镇)的社会治安状况是否满意?
□非常满意　□比较满意　□不太满意
□不满意

续表

14. 您对您所在村(镇)的班子或干部的管理工作是否满意?
□非常满意　□比较满意　□不太满意
□不满意
15. 您所在村(镇)有无下派大学生村官?
□有　□无
16. 您觉得大学生到了您村(镇),对您村(镇)的发展是否有所帮助?(15 题答案选择"有"的,填此题)
□大有帮助　□帮助不大
17. 您最希望大学生村官最能给你们提供哪方面的帮助?
□给我们带来科技致富的知识
□给我们带来好的发展点子和思路
□配合村干部搞好村里的工作
十二、您对国家在农村实行的各项惠农政策措施的态度
1. 您对贯彻"工业反哺农业,城市支持农村"的政策怎么看?
□非常好,确实落到了实处
□正在逐步落实,但实施起来困难很大
□基本上没有落实
□只停留在口头上
2. 您村(镇)对国家粮食综合补贴政策的贯彻执行:
□公开、公平、公正,农民满意,种粮积极性高涨
□存在不公正的现象,村干部克扣了一部分补贴
□存在不公平的现象,有些人没种那么多粮,却拿到了不少的补贴
□存在不公开的现象,种粮补贴,村里没有张榜公布
3. 您对移民搬迁工程是否满意?
□非常满意　□比较满意　□不太满意
□不满意
4. 您对提高农民素质培训工程是否满意?
□非常满意　□比较满意　□不太满意
□不满意
十三、您所在的村建设新农村的困难
1. 缺乏资金
您觉得缺乏资金的原因是:(可多选)
□集体经济薄弱,村里没有钱
□政府资金投入少
□老百姓有钱,不愿意拿
□老百姓没钱

续表

2. 农民致富难 您认为农民致富的主要困难是:(可多选) □缺乏资金　□缺少技术　□缺少劳动力 □土地太少　□信息不灵　□就业困难 □缺乏民间服务组织　□其他
3. 农民就业难 您觉得农民就业难的主要原因是:(可多选) □没有文化,很难找到合适的工作 □不愿意外出打工 □本地经济发展落后,不能提供充分的就业岗位
4. 发展现代农业难 您认为您所在的村(镇)发展现代农业的主要制约因素是:(最多可选三个答案) □耕地少,地力差　□小规模分散经营 □农业灌溉条件差　□农业物质技术装备机械化差 □农业适用技术推广难　□农业物资供应难 □农产品价格低且不稳定　□农产品市场需求信息不灵 □农民素质差,特别是留守农村劳动力素质低
5. 发展农村二、三产业难 您觉得发展农村二、三产业难的原因是:(可多选) □无资源　□无技术　□无人才 □无资金
6. 壮大村级集体经济难 您觉得壮大村级集体经济难的原因是:(可多选) □原来就没有基础,现在无从下手 □市场竞争激烈,要想壮大,不容易 □缺乏资金和人才 □缺乏资源
7. 农民素质提高难 您觉得农民素质提高难的原因有:(可多选) □农民认识不到知识的重要性 □职业教育发展滞后 □农民可以接受培训的机会太少了
8. 构建"工业反哺农业,城市扶持农村"长效机制难 您觉得构建"工业反哺农业,城市扶持农村"长效机制难的原因为:(可多选) □这是一项大工程,需要各方面的努力

续表

□政府在这方面的工作力度还不大 □社会对建设新农村的重要性还没有足够的认识，也没有形成共识 9．缩小城乡差距难 您觉得缩小城乡差距难的原因为：（最多可选三个答案） □城市各方面条件优越，经济发展越来越快，农村反之 □到农村投资的企业不多，城市反之 □城乡差距是长期历史形成的，差距是多方面的 □农村人口众多 □政府在加快农村发展方面的工作力度不够，措施还不多

您对新农村建设的期盼

十四、您认为您村开展新农村建设应从何处着手？（最多可选三个答案） □推进土地流转，实现规模经营 □发展现代农业 □发展农村工业 □促进农民增收 □加强基础建设，整治村容村貌 □因地制宜，搞好农村规划 □加强村级组织建设 □提高农民素质 □加强村级文化建设 □移民搬迁 □健全农村医疗卫生、养老、最低生活等各项保障制度 □加强对新农村建设工作的领导 □因地制宜，循序渐进 □要选一个好的带头人 □其他 十五、您对新农村建设有何期盼？（最多可选三个答案） □现有的“三农”政策稳定不变，未来的政策更加完善 □增加农业投入，搞好兴水沃土建设，提高农业综合生产能力 □加快小城镇建设，发展二、三产业，提供就业机会 □壮大集体经济，发展公益事业，改善生存环境 □增加社会事业投入，构建社会保障体系 □构建“工业反哺农业，城市支持农村”机制落在实处 □加强对新农村建设工作的领导

续表

十六、您希望通过哪些渠道实现早日致富？（最多可选三个答案）
□通过有偿流转获得更多的土地或养殖更多的牲畜，实现规模经营
□获得更多的技术或信息支持
□获得稳定的农字号龙头企业或专业协会的带动保障
□获得稳定的就业渠道
□通过参加各类培训提升自己的综合素质
□获得平等的国民待遇
十七、您认为应该如何发展现代农业？（最多可选三个答案）
□无论种植还是养殖要实现规模经营
□启动兴水沃土工程建设
□必须依靠科技进步，加快农业科技的推广应用步伐
□因地制宜发展特色农业
□有科技带头人的示范作用

附录 B　河北省农村老年人养老满意度调查

河北省农村老年人养老满意度调查问卷

您好，我们是河北经贸大学数学与统计学学院的学生，为了更好地了解社会，分析影响农村老年人养老的影响因素，探索合适的养老模式，我们组织了此次老年人养老满意度的调查。本次调查以研究为目的，采用无记名方式，请您放心作答。下面的问题可能会耽误您几分钟的时间，希望能得到您的支持，谢谢您的合作。

您的家庭住址：________省________市

1. 您的性别是：

□男　□女

2. 您的年龄是：

□50～60 岁　□60～70 岁　□70 岁以上

3. 您的文化水平是：

□没上过学　□小学　□初中

□高中　□大学及以上

4. 您所从事的职业：

□务农　□个体户　□外出务工

□公司职员　□公务员　□事业单位

□其他

5. 您的经济来源是：

□个人劳动所得　□养老保险　□离退休金

□政府补贴　□村集体补贴　□子女供养

□其他

6. 每月平均收入：

□100 元以下　□101～300 元　□301～500 元

□501～1000 元　□1001～2000 元　□2001～3000 元

□3000 元以上

续表

7. 储蓄：
□1000 元以下 □1001～5000 元 □5001～10 000 元
□10 001～50 000 元 □50 001～100 000 元 □100 001 元以上
8. 您的身体健康状况：
□非常健康 □健康 □一般
□体弱多病 □生活不能自理
9. 您的婚姻状况是：
□未婚 □已婚 □离异
□丧偶 □其他
10. 您的子女个数：________个；其中男孩个数为：________个。
11. 您的子女目前经常以什么方式给予您帮助或照料？
□经济上的支持 □日常生活照料 □精神上的依托
□其他
12. 您现在的居住状况是：
□独居或和配偶居住 □和子女居住 □养老机构居住
□其他
13. 您平时生病到哪里就医？
□县级以上医院 □乡卫生院
□村卫生室或私人诊所 □药店买药
□其他
14. 您有没有定期或不定期地体检？
□是 □否
如果有，是通过什么途径获得的？
□政府组织 □村集体组织
□医疗保健机构组织 □个人自愿体检
□其他
15. 近期您是否接触过老年人常见病、慢性病的预防健康指导？
□是 □否
如果有，是通过什么途径获得的？
□政府组织宣传 □村集体组织宣传
□医疗保健机构组织宣传
□电视报纸杂志宣传 □其他
16. 您对当地医疗服务机构的评价：
□非常满意 □满意 □一般
□不满意 □很不满意
17. 您居住所在地有没有村民娱乐活动场所？
□有 □没有

续表

如果有，您对娱乐活动场所的设施感觉如何？

□很满意　　□满意　　□一般

□不满意　　□很不满意

18. 您参加以下哪些社会活动(多选)？

□义务服务　　□社会捐助　　□老年书画活动

□老年文体活动　　□关心教育下一代　　□调解邻里纠纷和家庭矛盾

□维护社会治安　　□不参加　　□其他

每年参加社会活动的次数：

□0 次　　□1～3 次　　□4～6 次

□7～10 次　　□10 次以上

您对活动的满意程度：

□非常满意　　□满意　　□一般

□不满意　　□很不满意

19. 你是否有村集体的老年补贴？

□有　　□没有

若有村集体的老年补贴，每月的额度是多少？

□50 元以下　　□50～100 元　　□100～150 元

□150～200 元　　□200 元以上

20. 您是否有基本医疗保险或参加农村合作医疗？

□是　　□否

如果有，您对医疗保险或农村合作医疗的评价是：

□很好　　□好　　□一般

□不好　　□很不好

21. 您认为您所在村的低保制度落实情况如何？

□很好　　□好　　□一般

□不好　　□很不好

22. 您对老年生活的总体满意度如何？

□非常满意　　□满意　　□一般

□不满意　　□很不满意

参考文献

[1] Catania P. China's rural energy system and management[J]. Applied Energy, 1999(64): 229-240.

[2] Dincer I. Renewable energy and sustainable development: a crucial review [J]. Renewable and Sustainable Energy Reviews, 2000(4): 157-175.

[3] Johnson D E. Applied Multivariate Methods for Data Analysts[M]. Higher Education press,2005.

[4] Greenwood J, Jovanovic B. Financial Development Growth and the Distribution of Income[J]. Journal of Political Economy, 1990(98): 1076-1107.

[5] Lipton M. Why Poor People Stay Poor: Urban Bias in World Development [M]. London: Maurice Temple Smith Ltd, 1977.

[6] Qiu D X, Gu Sh H, Liang B F. Diffusion and Innovation in the Chinese Biogas Program[J]. World Development, 1990(18): 555-563.

[7] Rafael L P,Florencio L D S,Andrei S,Robert W V. Law and Finanee[J]. Journal of Political Economy,1998(6): 1113-1155.

[8] Ross L. Financial development and economic growth : views and agenda[J]. Journal of Economic Literature, 1997(35): 688-726.

[9] Ross L,Robert G K. Finance and Growth: Schumpeter Might Be Right[J]. The Quarterly Journal of Economics, 1993(108): 717-737.

[10] Tan P N, Michael S, Vipin K. Introduction to Data Mining [M]. Posts & Telecom press, 2006.

[11] Wamukonya L. Energy consumption in three rural Kenyan households: a survey[J]. Biomass and Bioenergy, 1995(6): 445-451.

[12] Wang R Y, Storey V C, Firth C P. A Framework for Analysis of Data Quality Research [J]. IEEE Trans Knowledge and Data Engineering, 1995 (7): 623-640.

[13] Wang X H, Feng Zh M. A Survey of rural energy consumption in the developed region of China[J]. Energy, 1997(5): 511-514.
[14] 爱德华·S. 肖. 经济发展中的金融深化(中文版)[M]. 上海:上海三联书店,1992.
[15] 雷蒙德·W. 戈德史密斯. 金融结构与金融发展[M]. 上海:上海人民出版社,1994.
[16] 胡锦涛在建设社会主义新农村专题研讨班上的重要讲话[N]. 人民日报,2006-2-16.
[17] 国务院. 中国老龄事业发展"十二五"规划. 2011.
[18] 国家统计局. 2010 年第六次全国人口普查主要数据公报(第 1 号). 2011.
[19] 中华人民共和国国民经济和社会发展第十一个五年规划纲要[N]. 人民日报,2006-3-12.
[20] 河北省人民政府关于推进社会主义新农村建设的指导意见. 冀发(2006)10 号.
[21] 河北省发展和改革委员会. 河北省"十一五"能源发展规划. http://www.hbdrc.gov.cn
[22] 河北省统计局. 河北经济年鉴[M]. 北京:中国统计出版社,2005-2012.
[23] 石家庄统计局. 石家庄统计年鉴[M]. 北京:中国统计出版社,2005-2010.
[24] 石家庄市政府. 石家庄市政府工作报告. http://www.sjz.gov.cn
[25] 沧州市政府. 沧州市政府工作报告. http://www.cangzhou.gov.cn
[26] 保定市政府. 保定市政府工作报告. http://www.bd.gov.cn
[27] 廊坊市政府. 廊坊市政府工作报告. http://www.lf.gov.cn
[28] 邢台市政府. 邢台市政府工作报告. http://www.xingtai.gov.cn
[29] 唐山市政府. 唐山市政府工作报告. http://www.tangshan.gov.cn
[30] 秦皇岛市政府. 秦皇岛市政府工作报告. http://www.qhd.gov.cn
[31] 张家口市政府. 张家口市政府工作报告. http://www.zjk.gov.cn
[32] 承德市政府. 承德市政府工作报告. http://www.chengde.gov.cn
[33] 衡水市政府. 衡水市政府工作报告. http://www.hengshui.gov.cn
[34] 邯郸市政府. 邯郸市政府工作报告. http://www.hd.gov.cn
[35] 中国人民大学统计学系数据挖掘中心. 统计学与数据挖掘[J]. 统计信息与论坛,2002(1): 4-9.
[36] 中国人民银行石家庄中心支行货币政策分析小组. 河北金融运行报告[M],2006-2010.

[37] 曹源芳. 中国区域金融中心体系研究[M]. 北京：中国金融出版社，2010.
[38] 陈立泰，黄仕川，李正彪. 金融深化对第三产业结构的影响分析——基于中国省际面板数据的分析[J]. 经济问题探索，2010(3)：76-79.
[39] 陈旭红，屈驳韵，李春林. 中国制造业竞争力的影响因素分析[J]. 河北经贸大学学报，2005(4)：49-54.
[40] 崔萍. 大连农村养老模式研究[D]. 大连：大连理工大学，2010.
[41] 戴林，李子奈. 农村能源综合建设项目社会经济效益及可推广性评价方法探讨[J]. 农业工程学报，2001(2)：115-118.
[42] 邓可蕴，等. 中国农村能源综合建设理论与实践[M]. 北京：中国环境科学出版社，2001.
[43] 邓可蕴. 以可再生能源服务农民 建设新农村[J]. 可再生能源，2007(2)：1.
[44] 董建军. 中国养老模式的社会化转型与社工介入[D]. 济南：山东大学，2010.
[45] 董金玲. 江苏区域金融作用机制及发展差异研究[D]. 北京：中国矿业大学，2009.
[46] 董少广，王淮海. 我国能源结构与资源利用效率分析[N]. 中国信息报，2006-4-19.
[47] 董绳周. 我国区域金融发展研究[D]. 天津：天津大学，2007.
[48] 杜宜宁. 辽宁区域金融与区域经济发展关系的实证分析[D]. 沈阳：辽宁大学，2011.
[49] 范方志，张立军. 中国地区金融结构转变与产业结构升级研究[J]. 金融研究，2003(11)：36-47.
[50] 方晓萍，等. 社会和谐程度的统计分析与评价[J]. 统计与决策，2007(24)：70-73.
[51] 费孝通. 费孝通论小城镇建设[M]. 北京：群言出版社，2000.
[52] 付瑶. 我国能源安全现状与对策[J]. 合作经济与科技，2007(7)：69-70.
[53] 高燕. 农村居民点用地整理的适宜性评价、模式及政策选择[D]. 杭州：浙江大学，2004.
[54] 关晓蕾. 基于决策树的分类算法研究[D]. 太原：山西大学，2006.
[55] 韩立民. 韩国的“新村运动”及其启示[J]. 中国农村观察，1996(4)：62-64.
[56] 韩明. 数据挖掘及其对统计学的挑战[J]. 统计研究，2001(8)：55-57.
[57] 杭雷鸣. 我国能源消费结构问题研究[D]. 上海：上海交通大学，2007.
[58] 和红. 我国区域金融协调发展问题研究[D]. 泉州：华侨大学，2007.

[59] 何芸，李建权.家庭结构变迁对养老模式的影响[J].社会工作，2007(1)：33-35.
[60] 胡伟.农民增收是当前农村能源建设工作的核心问题[J].农村能源，2000(3)：28-29.
[61] 黄湘雄.广东区域金融发展统计研究[D].长沙：湖南大学，2005.
[62] 金勇进，杜子芳，蒋妍.抽样技术[M].北京：中国人民大学出版社，2009.
[63] 李春林，等.区域产业竞争力——理论与实证[M].北京：冶金工业出版社，2004.
[64] 李放.农村老年人养老状况及其满意度的实证研究[J].开发研究，2010(1)：58-61.
[65] 李木祥，钟子明，冯宗茂.中国金融结构与经济发展[M].北京：中国金融出版社，2004.
[66] 李蕊.我国区域金融结构统计研究[D].长沙：湖南大学，2005.
[67] 李树琮.中国城市化与小城镇发展[M].北京：中国财政经济出版社，2002.
[68] 李新.中国金融发展与经济增长关系的研究[D].济南：山东大学，2009.
[69] 李振杰.草根调查——中国基层发展问题的社会学分析[M].北京：经济管理出版社，2004.
[70] 梁协雄，雷汝焕，曹长修.现代数据挖掘技术的研究进展[J].重庆大学学报，2004(3)：21-27.
[71] 林文文.我国金融发展与经济增长：基于联立方程计量模型的研究[D].宁波：宁波大学，2009.
[72] 刘斌，张兆刚，霍功.中国"三农"问题报告[M].北京：中国发展出版社，2004.
[73] 刘娟，黄午琴，崔霞.区域金融发展差异的综合评价[J].现代商贸工业，2012(20)：12-14.
[74] 刘孟芳.我国农村养老模式研究[D].武汉：武汉大学，2005.
[75] 刘薇.数据挖掘中决策树方法研究及其在房地产中介的应用[D].西安：西安电子科技大学，2006.
[76] 刘文倩，冯志胜.甘泉县农村生活能源结构调查及节柴途径的探讨[J].陕西林业科技，2002(3)：31-34.
[77] 刘勇.中国城镇化战略研究[M].北京：经济科学出版社，2004.
[78] 陆岭.数据挖掘在电信客户流失预测的应用[D].贵阳：贵州大学，2008.
[79] 麦金农.经济发展中的货币与资本(中文版)[M].上海：上海三联书

店,1993.
[80] 莫霞.农村可持续发展的人居环境建设研究[D].上海:同济大学,2006.
[81] 欧阳彬,吴茂群.基于因子分析的新农合农户满意度研究[J].农村经济与科技,2012(11):143-146.
[82] 彭旋子.基于农村居民意愿的养老模式选择研究[D].杭州:浙江农业大学,2010.
[83] 朴松花.黑龙江省经济发展与区域金融[D].吉林:吉林大学,2009.
[84] 冉茂盛.中国金融发展与经济增长作用机制研究[D].重庆:重庆大学,2003.
[85] 任彪.孙立辉.河北省金融发展与经济增长关系的统计研究[J].河北经贸大学学报,2011(12):71-74.
[86] 邵春妹.农村民居建设研究[D].石家庄:河北师范大学,2009.
[87] 沈坤荣,张成.金融发展与中国经济增长——基于跨地区动态数据的实证研究[J].管理世界,2004(7):15-21.
[88] 史丹.结构变动是影响我国能源消费的主要因素[J].中国工业经济,1999(11):38-43.
[89] 施骏.论发展沼气对农村可持续发展的贡献[J].安徽农业科学,2005(9):1785-1786.
[90] 师连枝.农村可再生能源发展与新农村建设[J].西南民族大学学报,2007(3):187-190.
[91] 苏胜强.CHAID分析在抽油烟机产品市场细分中的应用[J].商业研究,2006(17):105-110.
[92] 孙占潮.天津武清区农村沼气建设的发展[J].中国沼气,2006(3):58-59.
[93] 唐国新.我国区域金融差异及其空间计量分析[D].上海:复旦大学,2011.
[94] 唐志军.王玉霞."空心村"形成的深层次原因及其治理[J].中国发展观察,2008(3):56-57.
[95] 陶立群.人口城镇化过程中养老模式的创新与保障策略[J].上海城市管理,2010(4):83-85.
[96] 陶艺.中国宏观经济政策传导机制及政策效应的模拟分析[D].大连:东北财经大学,2005.
[97] 王海慧.河南省金融发展与经济增长关系的研究[D].成都:西南财经大学,2008.
[98] 王家城.赵志林.中国能源发展报告(2001)[M].北京:中国计量出版

社,2001.
[99] 王立成.推进城镇化 建设新农村[J].山东工商学院学报,2006(06):1-6.
[100] 王亮.中国地区金融发展与经济增长:panel data 分析[D].上海:华东师范大学,2007.
[101] 王明浩,陈佳骆,等.中国村镇发展中的住宅产业化研究[J].小城镇建设,2001(2):51-52.
[102] 王效华,冯祯民.农村能源可持续发展评价方法与实证[J].农业工程学报,2002(2):84-86.
[103] 王晓丽,师荣光.四川灾区农村重建中沼气生态农业技术开发利用研究[J].农业环境与发展,2008(4):58-61.
[104] 王旭.美国城市化的历史解读[M].长沙:岳麓书社,2003.
[105] 王兆星.中国金融结构论[M].北京:中国金融出版社,1991.
[106] 王正宽,吴争鸣.沼气池建设成本效益评价方法讨论[J].能源研究与利用,1999(4):48-49.
[107] 王治红,李蕾.河北与周边省份的金融发展比较研究[J].河北金融,2008(6):9-12.
[108] 武大勇.计量经济学中的面板数据模型分析[D].武汉:华中科技大学,2006.
[109] 本书编写组.建设社会主义新农村大参考[M].北京:红旗出版社,2006.
[110] 奚维吉.用户满意度调查的数据处理[J].科技资讯,2007(8):253-254.
[111] 项利华.我国区域金融发展研究[D].哈尔滨:哈尔滨工程大学,2006.
[112] 肖云,刘培森.新型农村社会养老保险满意度影响因素分析[J].经济体制改革,2011(5):66-70.
[113] 谢治国.新中国能源政策研究[D].北京:中国科学技术大学,2006.
[114] 薛力.城市化背景下的“空心村”现象及其对策探讨——以江苏省为例[J].城市规划,2001(6):9-14.
[115] 薛薇,陈欢歌.数据挖掘方法及应用[M].北京:电子工业出版社,2012.
[116] 闫海成.论我国的能源发展战略[D].长春:东北师范大学,2006.
[117] 杨断涛.滇西北老君山地区农村能源及其产业发展研究[D].郑州:河南农业大学,2004.
[118] 杨宏.人口老龄化形势下发展中国老龄产业[J].大连海事大学学报,2006(2):1-2.
[119] 杨洪涛,李桂君.关联规则在房地产广告媒体选择中的应用[J].计算机工

程与应用,2006(5):230-232.
[120] 杨家栋,秦兴方,单宜虎.农村城镇化与生态安全[M].北京:社会科学文献出版社,2005.
[121] 杨伟.河北区域金融发展差异研究[D].石家庄:河北经贸大学,2011.
[122] 姚晨光.河北省水资源开发利用现状与对策[J].海河水利,2006(4):6-7.
[123] 殷德生,肖顺喜.体制转轨中的区域金融研究.[M]上海:学林出版社,2000.
[124] 尹优平.中国区域金融协调发展研究[D].成都:西南财经大学,2007.
[125] 于波.农户能源选择因素分析及对新能源推广的政策含义分析[D].南京:南京农业大学,2007.
[126] 于秀林,任雪松.多元统计分析[M].北京:中国统计出版社,1999.
[127] 曾铮,胡志浩,曾华珑.中国的金融发展与经济增长:一个联立方程计量模型的经验解释[J].金融发展研究,2008(7):25-29.
[128] 张春姹.河北区域金融差异实证研究[D].石家庄:河北经贸大学,2011.
[129] 张胆.农村居民养老模式选择意愿及影响因素分析[D].福州:福建农林大学,2009.
[130] 张光波,刘复国.增加农民收入的关键在于提高农民素质[J].潍坊学院学报,2004(3):131-133.
[131] 张军洲.中国区域金融分析[M].北京:中国经济出版社,1995.
[132] 张丽娜.重庆市农村能源供需分析及其政策研究[D].重庆:重庆大学,2006.
[133] 张尧庭,谢邦昌,朱世武.数据采掘入门及其应用[M].北京:中国统计出版社,2001.
[134] 张巍.黑龙江省农村社区养老模式研究[D].西安:陕西师范大学,2012.
[135] 章文波,陈红艳.实用数据统计分析及 SPSS 12.0 应用[M].北京:人民邮电出版社,2006.
[136] 张铁倩.河北省区域金融发展问题思考[D].保定:河北大学,2010.
[137] 张湛,刘劲. 建设公寓式住宅——改变农村散居现状,促进农村经济发展[J].小城镇建设,1999(12):27.
[138] 赵俊俊.我国农村劳动力转移背景下养老模式研究[D].北京:首都师范大学,2009.
[139] 赵秋立.浙江省农村人居环境优化建设研究[D]. 杭州:浙江大学,2006.
[140] 赵艳霞,王立东,尹景瑞.在乡村治理的路径分析——基于唐山农村新民

居建设的调查[J].前沿,2009(12):174-176.
[141] 赵之枫.城市化加速时期村庄集聚及规划建设研究[D].北京:清华大学,2001.
[142] 周立,王子明.中国各地区金融发展与经济增长实证分析1978—2000[J].金融研究,2002(10):1-13.
[143] 朱建芳.区域金融发展差距:理论与实证分析[D].杭州:浙江大学,2006.
[144] 朱建平,谢邦昌.数据挖掘中关联规则的提升及其应用[J].统计研究,2004(12):34-39.
[145] 朱向东.农村经济调查研究文集[M].北京:中国统计出版社,2000.
[146] 朱亚男.不平衡发展条件下的农村养老模式选择[D].保定:河北农业大学,2008.
[147] 资树荣.发展农村沼气产业 建设节约型新农村[J].理论学刊,2006(1):49-50.

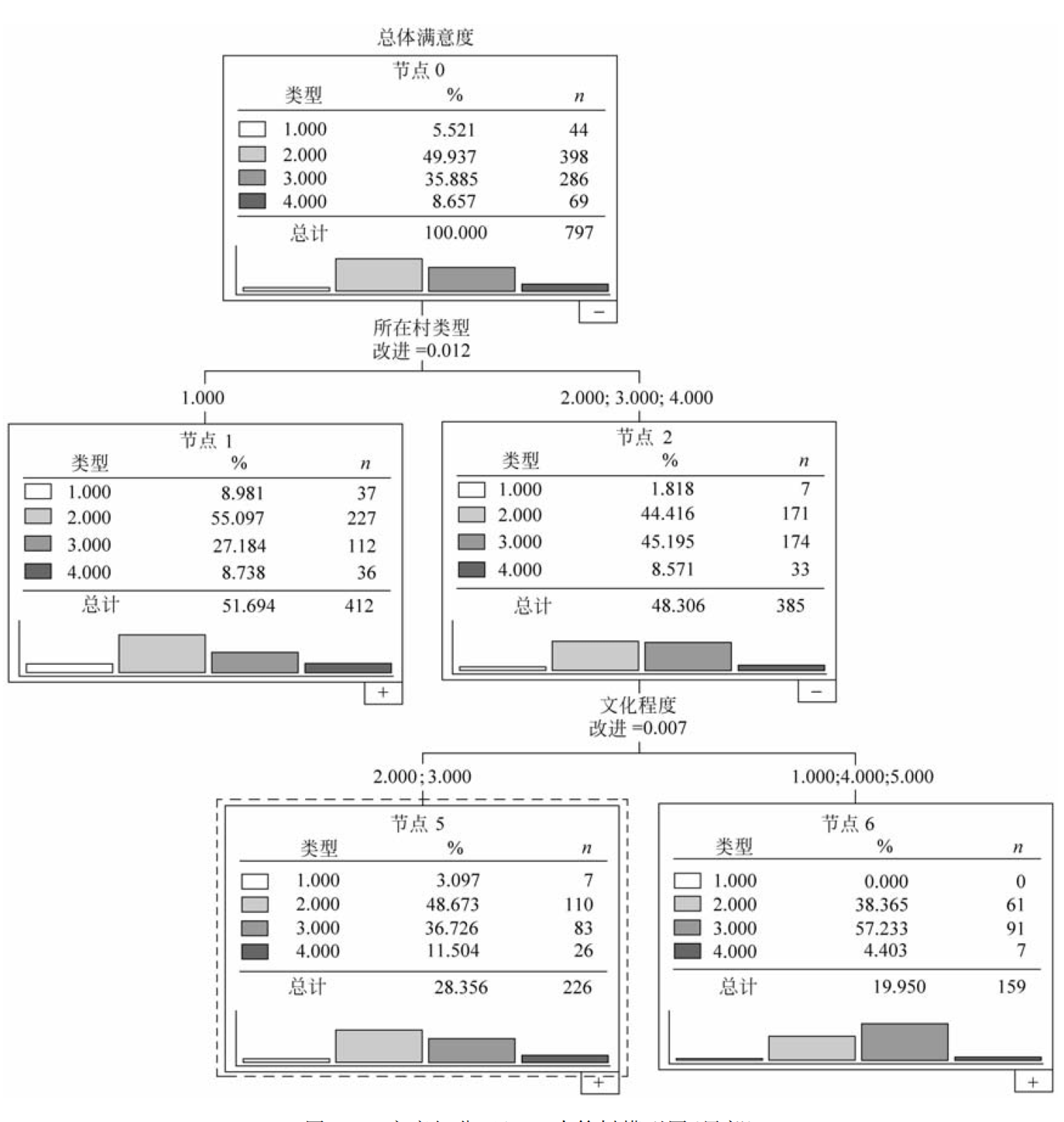


图 9-3　广度细分 CART 决策树模型图(局部)

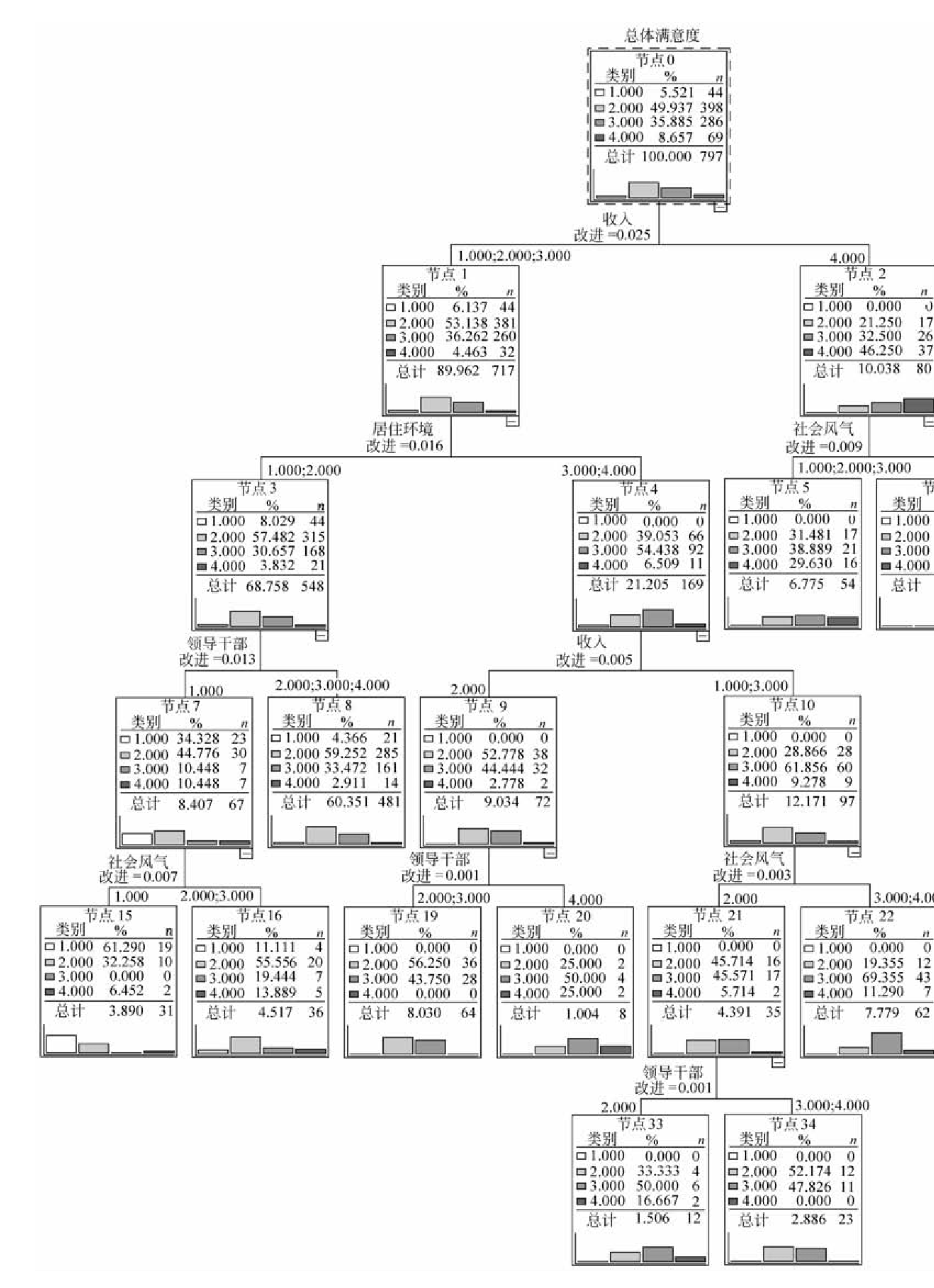


图 9-5　深度细分 CART 决策树模型图(局部)

4.000

点 6

%	n
0.000	0
0.000	0
19.231	5
80.769	21
3.262	26

00

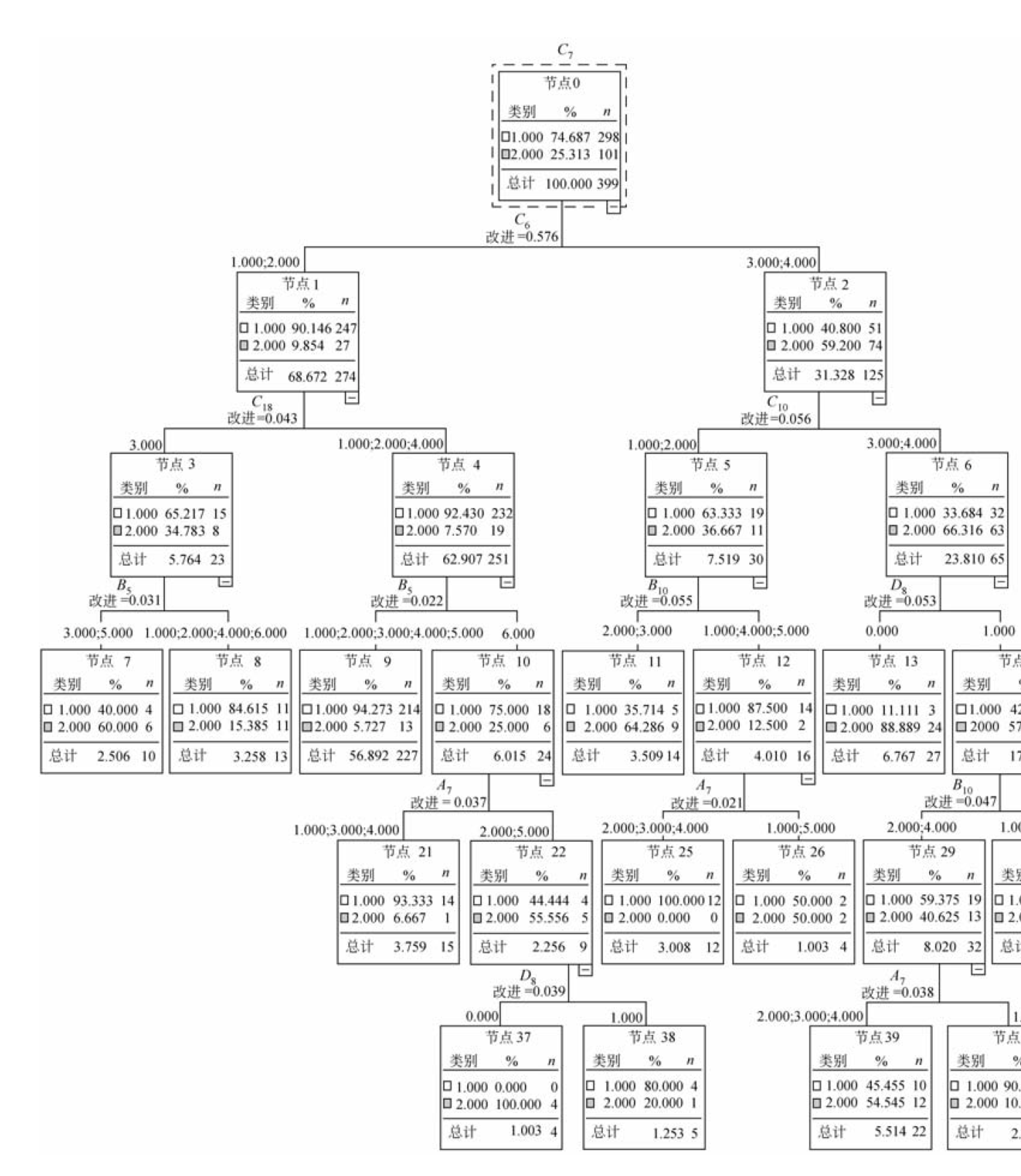


图 12-2　农村生活现代化对农民生活满意度的决策树

点 14

% n

2.647 29
.353 39

.043 68

00;3.000;5.000

节点30

别 % n

000 27.778 10
000 72.222 26

计 9.023 36

.000;5.000

40

% n

000 9
000 1

506 10

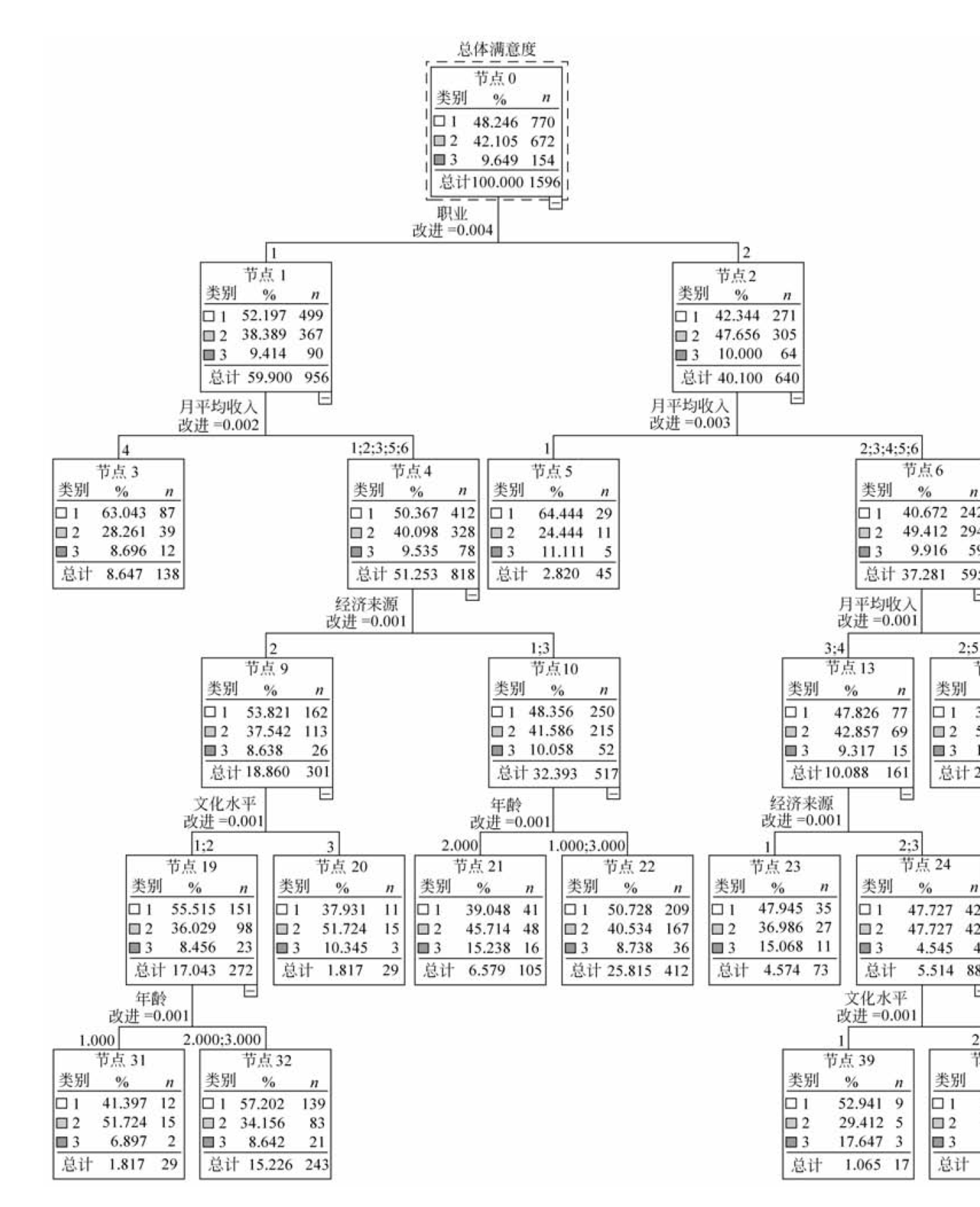


图 18-3　广度细分分类回归树模型

:6

节点 14	
%	*n*
38.018	165
51.843	225
10.138	44
7.193	434

:3

节点 40	
%	*n*
46.479	33
52.113	37
1.408	1
4.449	71

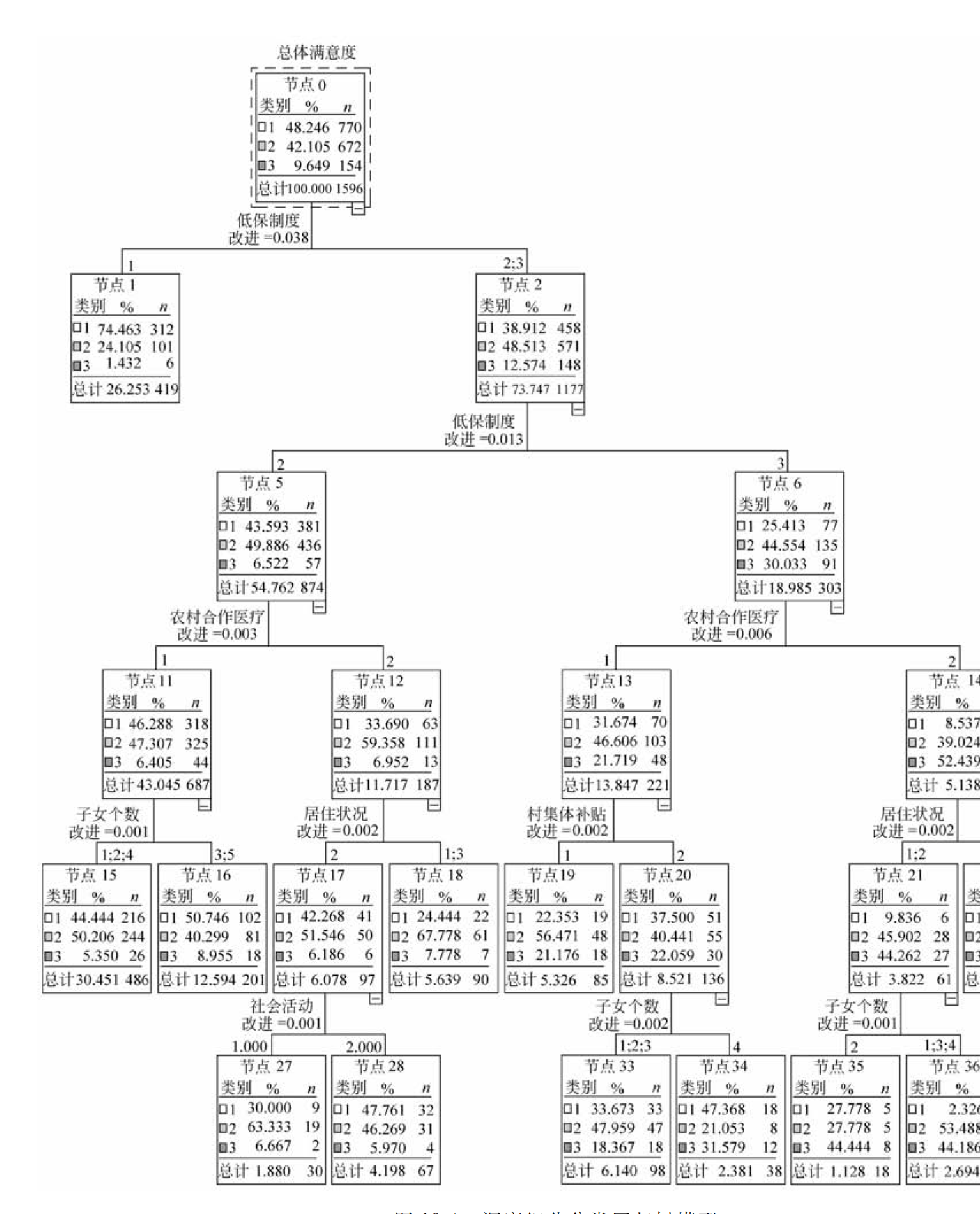


图 18-4　深度细分分类回归树模型

n
7
32
43
82

3

节点22

别	%	n
	4.762	1
	19.048	4
	76.190	16
计	1.316	21

n
1
23
19
43